SCHOLARLY PUBLICATIONS

A CURRENT AWARENESS BULLETIN
OF RESEARCH OUTPUT

@DTU

(113th Edition)

MAY 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 Thirteenth** 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 **May, 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 examination of colorectal cancer burden in European Union, 1990–2019: Estimates from Global Burden of Disease 2019 Study, ***3.Rajesh Sharma***, USME, DTU

2. A Comparative Study of Data-driven Models for Groundwater Level Forecasting, ***6.R. Sarma*** and ***3.S. K. Singh***, Environmental, DTU

3. A Comprehensive Study on the Role of Machine Learning in Hybrid Biometric Recognition, ***7.Shipra Shand*** and ***3.Rahul***, Software Engineering, DTU

4. A Deep Learning Approach for Interior Designing of Apartment Building Architecture using U2 Net, ***8.Gaurav Kumar***, ***8.Ashish Kumar*** and ***3.Rahul***, Software Engineering, DTU

5. A novel authentication scheme for group based communication for IoT oriented infrastructure in smart cities, ***6.Irfan Alam*** and ***3.Manoj Kumar***, CSE, DTU

6. A Novel CFDITA-Based Design of Grounded Capacitance Multiplier and Its Transpose Structure, ***8.Atul Kumar***, ***8.Dheeraj Singh*** and ***3.Deva Nand***, ECE, DTU

7. A novel intelligent strategy–based thermodynamic modeling and analysis of solar-assisted vapor absorption refrigeration system, Harsh Kadyan, Anil Kumar Berwal and ***3.Radhey Shyam Mishra***, Mechanical, DTU

8. A REVIEW OF THE DRIP IRRIGATION SYSTEM, ***8.Bhagya Jayant***, ***8.Kunal Dahiya***, ***8.Ark Rukhiyar***, ***3..Ritu Raj*** and ***7.Rahul Kumar Meena***, Civil, DTU

9. A Structural Approach on Enablers of IoT for Sustainable Development of Smart Cities, ***8.Ayusha Thakur***, Satyabrata Aich and ***3.Rohit Kumar***, Electronics, DTU

10. A STUDY ON CLOUD MANUFACTURING, ***3.Dr. Ak Madan***, ***6.Bharath Gohrani***, ***8.Akshat Singh*** and ***8.Adarsh Pillai***, Mechanical, DTU

11. A Systematic Review on Breast Cancer Detection Using Deep Learning Techniques, ***6.Kamakshi Rautela***, ***3.Dinesh Kumar*** and Vijay Kumar, Electroncis, DTU

12. An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete, Arti Chouksey, ***6.Manvendra Verma***, ***3.Nirendra Dev***, Ibadur Rahman and Kamal Upreti, Civil, DTU

13. Analog/RF Performance and Effect of Temperature on Ferroelectric Layer Improved FET device with Spacer, ***6.Yash Pathak***, ***3.Bansi Dhar Malhotra*** and ***3.Rishu Chaujar***, Applied Physics and Biotechnology, DTU

14. Analysis of melting of some metals using pressure dependence of the Grüneisen parameter in the Lindemann law, D. Ashwini, ***Vijay S. Sharma*** and K. Sunil, Mechanical, DTU

15. Analysis of wind turbine dataset and machine learning based forecasting in SCADA-system, ***6.Upma Singh*** and ***3.M. Rizwan***, Electrical, DTU

16. Carbon Dioxide Filtering Module, ***8.Annoy Roy*** & ***8.Shivani Kumari***, Chemistry, DTU

17. Comparative Analysis of Time Series Models on COVID-19 Predictions, ***8.Puneet Kumar Sehrawat*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

18. Comparative Analysis of Transformer Models on WikiHow Dataset, ***8.Dev Jadeja***, ***8.Avinash Khetri***, ***8.Anubhav Mittal*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

19. Computer-controlled diabetes disease diagnosis technique based on fuzzy inference structure for insulin-dependent patients, ***6.Ankit Sharma***, ***3.Nilam*** and Harendra Pal Singh, Applied Mathematics, DTU

20. Deep Learning Methods for Anomaly Detection, Rajeev Subedi, ***7.1.Anil Kumar*** and Narendra Kumar, CSE, DTU

21. Design and Analysis of 55–63-GHz Fundamental Quad-Core VCO With NMOS-Only Stacked Oscillator in 28-nm CMOS, ***8.1.Sriram Balamurali***, Giovanni Mangraviti, Cheng-Hsueh Tsai, Piet Wambacq and Jan Craninckx, Electrical, DTU

22. Design and analysis of Quadratic Bernstein Functional Blending Neural Network for shunt compensation and Phase Locked Loop, ***3.Ankita Arora*** and ***3.Alka Singh***, Electrical, DTU

23. DESIGN AND ANALYSIS OF SINGLE POINT CUTTING TOOL, ***3.A.K. Madan***, ***8.Badal Singh***, ***8.Devansh Yadav*** and ***8.Deepak Kumar***, Mechanical, DTU

24. DESIGN AND MODELLING OF SQUARE HOLE DRILLING MACHINE, Dr. ***3.Ashok Kumar Madan***, ***8.Chaudhary Jatin***, ***8.Chiranjeev Singh Sethi*** and ***8.Deepak Kumar***, Mechanical, DTU

25. Design of CDTA and VDTA Based Frequency Agile Filters, ***3.Neeta Pandey***, ***8.Aseem Saya***, ***8.Richa Choudhary*** and ***3.Rajeshwari Pandey***, Electronics and Electrical, DTU

26. Detection of biomolecules in dielectric modulated double metal below ferroelectric layer FET with improved sensitivity, ***6.Yash Pathak***, ***3.Bansi Dhar Malhotra*** and ***3.Rishu Chaujar***, Applied Physics and Biotechnology, DTU

27. Detection of Defective Steel Surface with Image Segmentation, ***8.Chirag Gupta***, ***7.1.Gaurav Chauhan*** and ***8.Hari Dev S Aiyar***, Mechanical, DTU

28. Disease-associated dysbiosis and potential therapeutic role of Akkermansia muciniphila, a mucus degrading bacteria of gut microbiome, ***8.Vidushi Aggarwal***, ***8.Sushant Sunder*** and ***3.Smita Rastogi Verma***, Biotechnology, DTU

29. Earth-air heat exchanger and its utilization in cold storages, ***8.Palak Kumari***, ***8.Martin Leo King*** and ***3.Dr. JP Kesari***, Mechanical, DTU

30. Effect of leaves and seeds of Achyranthes aspera as feed supplements on the immunological and stress parameters and related gene expressions of Asian catfish (Clarias batrachus), ***3.Neelesh Kumar***, ***3.JaiGopal Sharma***, ***6.Prabhat Mittal*** and ***6.Rina Chakrabarti***, Biotechnology, DTU

31. Effect of Wind Loads on Low-Rise Pitched and Circular Arch Roof Structures: A Comparative Study based on Numerical Simulation, ***8.Aditya Kumar Jha***, ***8.Amartya Sinha*** and ***3.Ritu Raj***, Civil, DTU

32. Electrochemical performance of NCA based cathodes for variable thickness of electrode through modelling and simulation, ***6.Pritesh Singh***, ***7.Vaishali Dudeja***, ***3.Amrish K. Panwar***, Applied Physics, DTU

33. Enhancement of Microstructural and Mechanical Properties of Ultrasonic Vibration-Assisted Cold Metal Transfer Welding of 304 Stainless Steel, ***3.N. Yuvaraj***, ***6.Jayanta Ghosh Roy***, and ***3.Vipin***, Mechanical, DTU

34. Enhancing the Thermal Performance of a Double Pipe Heat Exchanger in Turbulent Flow Conditions, ***6.Manish Sanserwal***, Devendra Yadav and Mayank Bhardwaj and Gurjeet Singh, Mechanical, DTU

35. Ethical Hacking of IEEE 802.11 Encryption Protocols, ***8.Sandesh Jain***, ***8.Sarthak Pruthi*** and ***8.Vivek Yadav***, IT, DTU

36. Evaluation of Phytocompounds of Vitis vinifera, Tabernaemontana corymbosa and Aloe vera against JAB1: A Strategy to Treat Cervical Cancer, ***8.Ritika Luthra***, ***8.Nikita Sharma***, ***8.Shreeja Datta*** and ***3.Asmita Das***, Biotechnology, DTU

37. Exploring intergenerational differences in the virtue of appreciation at the workplace, ***3.Naval Garg*** and Manju Mahipalan, USME, DTU

38. Fetal state health monitoring using novel Enhanced Binary Bat Algorithm, ***8.Prerna Sharma*** and ***3.Kapil Sharma***, IT, DTU

39. Finding patterns in Knowledge Attribution for Transformers, ***8.Jeevesh Juneja*** and ***3.Ritu Agarwal***, IT, DTU

40. FLOATING SOLAR PHOTOVOLTAIC SYSTEM, ***8.Ayush Goel***, ***8.Deepanshu Yadav*** and ***3.J.P. Kesari***, Mechanical, DTU

41. FRUITS/VEGETABLE CLASSIFICATION USING DEEP LEARNING AND IOT, ***3.Ashish Rajeshwar Kulkarni***, ***8.Amit Gupta***, ***8.Ajit Kumar Yadav*** and ***3.Anand Kumar***, Electrical, DTU

42. Green route synthesis and characterization of zinc oxide nanorods and its composite with PVA, ***7.Parisha Diwan***, ***6.Kanika Sharma***, ***3.Nitin K. Puri***, Applied Physics, DTU

43. Highly Sensitive Enzymatic Biosensor Based on Polyaniline-Wrapped Titanium Dioxide Nanohybrid for Fish Freshness Detection, ***6.Deeksha Thakur***, ***3.Chandra Mouli Pandey*** and ***3.D. Kumar***, Applied Chemistry, DTU

44. Hydro-Electric Power Generation, ***8.Mukul Chauhan***, ***8.Neeraj Joshi*** and ***3.JP Kesari***, Software Engineering and Mechanical, DTU

45. Image Compressed Sensing Using Deep Learning, ***8.Saksham Gera***, ***8.Shiva Gupta***, ***8.Paras HS***, ***3.Rohan Pillai*** and ***3.Madan Mohan Rayguru***, Electrical, DTU

46. Implementation of mass customization for competitive advantage in Indian industries: an empirical investigation, ***6.Piu Jain***, ***3.Suresh Garg*** & Gayatri Kansal, DTU

47. Integrating node centralities, similarity measures, andmachine learning classifiers for link prediction, Sameer Anand, ***3.Rahul***, ***8.Abhishek Mallik*** & ***3.Sanjay Kumar***, CSE.

48. Interrelation Between Gratitude and Vitality Among Students: Exploring Mediating Role of Resilience, ***3.Naval Garg***, Sarika Kumari and B. K. Punia, USME, DTU

49. Investigating Exhaust Emissions from In-Use Passenger Cars: Exploratory Analysis and Policy Outlook, ***6.Abhinav Pandey***, ***3.Rajeev Kumar Mishra*** and Govind Pandey, Environmental, DTU

50. Investigation of Circularly Polarized CPW fed Antenna as a 2.45 GHz RFID Reader, Amit Birwal, Vipul Kaushal and ***6.1.Kamlesh Patel***, Electronics, DTU

51. Investigation of electronic, mechanical and thermoelectric properties of quaternary Heusler compounds ZrRhTiZ (Z = ln, Al), ***7.Vivek Kumar***, ***6.Mohit Kumar*** and ***3.Mukhtiyar Singh***, Applied Physics, DTU

52. Investigations on effect of bending radius on springback behaviour of three-ply clad sheet, ***6.Pankaj Kumar Sharma***, ***3.Vijay Gautam*** and ***3.Atul Kumar Agrawal***, Mechanical, DTU

53. IOT BASED SANITIZATION ROBOT USING RASPBERRY PI ZERO W AND UV LAMP, ***8.B Revanth Raj***, ***8.Abhishek Sahani*** and ***8.Ravi Yadav***, Mechanical, DTU

54. Key role of Tb3+ doping on structural and photoluminescence properties of Gd2Ti2O7 pyrochlore oxide, ***6.Umang Berwal***, ***3.Vinod Singh*** and ***3.Rinku Sharma***, Applied Physics, DTU

55. Linear Connectivity Reveals Generalization Strategies, ***8.Jeevesh Juneja***, ***8.Rachit Bansal***, Kyunghyun Cho, Jo˜ao Sedoc and Naomi Saphra, DTU

56. LOW THD GRID-CONNECTED INVERTER FOR TWO PV ARRAY, ***3.Vinay***, ***8.Parag Agrawal*** and ***8.Neeraj Kumar***, DTU

57. Manufacturing of Regeneratively Cooled Rocket Nozzles, ***3.Dr. A.K. Madan***, ***8.Prakash Narayan Budhaulia*** and ***8.Priyanshu Nailwal***, Mechanical, DTU

58. Mapping incidence and mortality of leukemia and its subtypes in 21 world regions in last three decades and projections to 2030, ***3.Rajesh Sharma*** and ***6.Chinmay Jani***, USME, DTU

59. Measurement of Tribological Characteristics of Composite Coating Under Different Temperatures, ***6.K. Gupta***, ***3.Q. Murtaza*** and ***3.N. Yuvraj***, Mechanical, DTU

60. Modelling and Inverse Kinematics of Tendon Driven Continuum Robots using Cascade Correlation, ***8.Ashok Madan***, ***8.Shaswat Garg*** and ***8.Satwik Dudeja***, Mechanical, DTU

61. Multi – Objective Parametric Optimization of Milling on AMMC using Supervised Machine Learning Techniques, ***3.Dr. A.K. Madan***, ***8.Pratyush Bhatt*** and ***8.Pranav Taneja***, Mechanical, DTU

62. MULTI-FEATURE AWARE POSE AND GEOMETRY BASED FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING, ***6.Ankit Sharma*** and ***3.Dr. Manoj Sethi***, CSE, DTU

63. Multimodal and Multilabel Genre Classification Of Movie Trailers, ***8.Aditya Kishore Jha***, ***8.Akshat Batra***, ***8.Akshat Dubey*** and ***3.Dinesh Kumar Vishwakarma***, IT, DTU

64. My Land My Will - A Novel Blockchain Based Land Registry System, ***8.Rohan Vashist***, ***8.Saumya Swarup***, ***8.Sneha Sood***, ***3.Anamika Chauhan*** and ***3.Anupama***, Electrical and IT, DTU

65. Numerical investigation and temperature-based analysis of the analog performance of fully gate-covered junctionless FinFET, ***8.Gaurav Mangal***, ***8.Aman Tyagi*** and ***3.Rishu Chaujar***, Applied Physics, DTU

66. Numerical study of 3-D helical passive micromixer having both inlets at offset with blood as fluid, ***7.Akash Sinha*** and ***3.M. Zunaid***, Mechanical, DTU

67. Performance Evaluation & Linearity Distortion Analysis for Plasma-Assisted Dual-Material Carbon Nanotube Field Effect Transistor with a SiO2-HfO2 Stacked Gate-Oxide Structure (DM-SGCNFET), ***6.Mansha Kansal*** and ***3.Suresh C. Sharma***, Applied Physics, DTU

68. Potential and Achievement of Renewable Energy sources in India, ***8.Ayush Bansal***, ***8.Ayush Jha*** and ***3.J. P. Kesari***, Applied Physics and Mechanical, DTU

69. Potential of a Nonperennial Tributary Integrated with Solar Energy for Rural Electrification: A Case Study of Ikukwa Village in Tanzania, ***8.Isaka J Mwakitalima***, ***3.Mohammad Rizwan*** and ***3.Narendra Kumar***, Electrical, DTU

70. Predicting defects in imbalanced data using resampling methods: an empirical investigation, ***3.Ruchika Malhotra*** and ***6.Juhi Jain***, Software Engineering and CSE.

71. PREDICTION OF CUTTING ENERGY AND FORCES ENCOUNTERED IN TURNING PROCESS USING CAD SOFTWARE, ***3.A.K. Madan***, ***8.Aditya Kumar***, ***8.Yash Kumar***, Mechanical, DTU

72. PROCEEDINGS OF ITR INTERNATIONAL CONFERENCE Kolkata, India Organized by Date of Event, ***7.Debottam Bhowmik***, Production, DTU

73. Process optimization of conventional steam distillation system for peppermint oil extraction, ***8.1.Ravi Kant*** and ***3.Anil Kumar***, Mechanical and Environment, DTU

74. Ranking Users in the Social Network: A Survey of Budget and Opinion dependent Methods, ***8.Akshit Tomar***, ***8.Aman Dhiman*** and ***8.Abhishek Midha***, CSE, DTU

75. Reweighted L0 norm variable step size continuous mixed p-norm control scheme for mitigating power quality problems of grid coupled solar PV system, Pallavi Verma, ***6.Avdhesh Kumar***, Rachana Garg and ***3.Priya Mahajan***, Electrical, DTU

76. Semantic segmentation in medical images through transfused convolution and transformer networks, ***8.Tashvik Dhamija***, ***8.Anunay Gupta***, ***8.Shreyansh Gupta***, ***6.Anjum***, ***3.Rahul Katarya*** and Ghanshyam Singh, ECE, Electrical, Civil & CSE.

77. Solar Energy: A source for E-vehicles, ***6.Kaseri Abdul Rehman***, ***3.Naveen Kumar*** and ***3.JP Kesari***, Mechanical, DTU

78. Solar Flare Prediction using Machine Learning Algorithms on RHESSI Dataset, ***8.Moulshree***, ***8.Munira Anjum*** and ***3.Anurag Goel***, CSE, DTU

79. Spectroscopic Studies of Eu3+ Ion-Doped Antimony-Lead-Oxyfluoroborate Glasses for Visible Red Photonic Device Applications, D. Shanmukha Sai, K. Swapna, K. Deepthi, J. V. Shanmukh Kumar, Sk. Mahamuda, M. Venkateswarlu, P. Sruthi, D. Amer and ***3.A. S. Rao***, Applied Physics, DTU

80. Static analysis of multi-metal usage in grid stiffened structures, ***8.Kshitij Tripathi***, ***8.Kunal Kukreja*** and ***3.Dr Vikas Rastogi***, Mechanical, DTU

81. Structural and spectroscopic studies of Eu3+ activated potassium bismuth molybdate phosphor for optoelectronic device applications, ***7.Divyanshi Nagpal***, ***7.Tanisha Bhadauria***, ***3.M. Jayasimhadri***, Applied Physics, DTU

82. Structural matrices for Signed Petri net, ***3.Payal*** and ***3.Sangita Kansal***, Applied Mathematics, DTU

83. STUDY ABOUT MACHINE OPERATIONS AND ITS DIFFERENT OPERATIONS, ***8.Chander Mohan Aggarwal***, ***8.Bhavya Dhanwaria***, ***3.Dr. Ravi Butola*** and ***3.Prof. A.K. Madan***, Mechanical, DTU

84. Surface Roughness in Abrasive Mixed Rotary Electrical Discharge Machining: Experimental Approach, A. Lamba and ***3.V. Vipin***, Mechanical, DTU

85. Systematic study of topology control methods and routing techniques in wireless sensor networks, ***6.Radhika Kavra***, ***3.Anjana Gupta*** and ***3.Sangita Kansal***, Applied Mathematics, DTU

86. TLMOTE: A Topic-based Language Modelling Approach for Text Oversampling, ***8.Arjun Choudhry***, ***3.Seba Susan***, ***8.Anmol Bansal*** and ***8.Anubhav Sharma***, IT, DTU

87. v-TCM: Vertical-aware Transformer Click Model for Web Search, ***7.Kritarth Bisht*** and ***3.Seba Susan***, IT, DTU

88. Weight Optimization and Structural Analysis of an Electric Bus Chassis Frame, ***8.Rahul Saini***, ***8.Rishav Jindal***, ***8.Sandeep Kumar***, ***3.Mohd.Tayyab*** and ***3.Ranganath M. Singari***, Mechanical, Design, DTU

89. WIND INTERFERENCE ON A HEXAGONAL SHAPED HIGHRISE BUILDINGS WITH DIFFERENT OPENINGS, ***8.Harsh Goyal***, ***8.Harmanbir Singh***, ***8.Mayank Singh*** and ***3.Ritu Raj***, Civil, DTU

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

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

International Journal of Clinical Oncology
https://doi.org/10.1007/s10147-022-02182-0

ORIGINAL ARTICLE

A comparative examination of colorectal cancer burden in European Union, 1990–2019: Estimates from Global Burden of Disease 2019 Study

Rajesh Sharma[1]

Received: 2 February 2022 / Accepted: 25 April 2022
© The Author(s) under exclusive licence to Japan Society of Clinical Oncology 2022

Abstract

Aim This study examines the burden of colorectal cancer (CRC) in European Union (EU) countries in the last 3 decades.

Methods The data pertaining to CRC burden were procured from the Global Burden of Disease 2019 Study for 28 EU countries (including United Kingdom) for the period 1990–2019. The age-standardized rates of CRC were utilized to compare the country-wise burden and joinpoint regression models were applied to examine the trends.

Results In EU, CRC incident cases increased by 70.2% from 261,306 to 444,872 and deaths increased by 36.8% from 155,823 to 213,174 between 1990 and 2019. The age-standardized incidence rate (ASIR) increased by 11.9% from 37.8/100,000 to 42.3/100,000 between 1990 and 2019; in contrast, the age-standardized mortality rate (ASMR) decreased by 16.9% (1990: 22.4/100,000; 2019: 18.6/100,000) and age-standardized DALYs rate (ASDALR) decreased by 18.6% (1990: 472.9/100,000; 2019: 385.1/100,000) in the study period. In 2019, Hungary was the leading country in terms of ASMR (28.6/100,000) and ASDALR (630.3/100,000), and Lithuania (29.2/100,000) had the lowest ASIR, whereas Finland had the lowest ASMR (12.3/100,000) and ASDALR (253.6/100,000) in 2019.

Conclusion CRC incidence is increasing in EU and mortality rates, although decreasing, are still unacceptably high. CRC control efforts must be focused around early detection using screening and prevention through reduction of modifiable risk factors. Increasing CRC incidence rates in young adults in recent years requires more research to pinpoint risk factors, and there must be more awareness of this recent development among general public and clinicians.

Keywords Colorectal cancer · Incidence · Mortality · Disability-adjusted life years · Global Burden of Disease · European Union

Introduction

Colorectal cancer (CRC) is one of the leading cancers worldwide, resulting in 1.9 million new cases and 935,173 deaths in 2020, making it the second leading cause of cancer deaths worldwide [1]. Behavioral risk factors, including diet, high body mass index (BMI), smoking, and less physical activity, account for large percentage of new CRC cases [2]. As per World Cancer Research Fund, consuming wholegrains, dairy products, calcium supplements, and dietary fiber have protective effects against weight gain, obesity and CRC, while consuming red and processed meat increase the risk of CRC [3]. About 10–15% of CRC is familial (lynch syndrome and familial adenomatous polyps), out of which 5% is linked with highly penetrant dominant inherited syndrome [4]. Lynch syndrome is the most common form of inherited CRC, accounting for 2–4% of all CRC cases [5].

CRC is the second leading cancer in Europe, with 519,820 incident cases and 244,824 deaths in 2020, with Europe accounting for 26.1% of CRC deaths and 26.9% of global incident cases [1]. Among 21 world regions, Australia/New Zealand had the highest age-standardized incidence rate (ASIR: 447.2/100,000) with all European regions (western, northern, southern, central, and eastern) had ASIR higher than global ASIR of 201.0/100,000. The total economic cost of CRC to Europe was €19.1 billion in 2015

✉ Rajesh Sharma
rajesh.sharma@dtu.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, East Delhi Campus, Room No. 305, Vivek Vihar Phase II, Delhi 110095, India

Published online: 20 May 2022

[6]. Based on an estimate [4], 3 million people aged 45 to 70 years are at risk of familial CRC in Europe.

Many deaths from CRC can be prevented through early detection of CRC, which can be treated, leading to better survival rates. In this endeavor, screening through stool-based tests or invasive procedures such as flexible sigmoidoscopy and colonoscopy has been implemented in different countries as organized or opportunistic screening tools. In 2003, the European Council recommended organized CRC screening among individuals aged 50–74 years using guaiac fecal occult blood test (gFOBT) [7]. In 2012, these recommendations were updated to include Fecal immunochemical test (FIT), flexible sigmoidoscopy (FS), and colonoscopy as reasonable alternatives of gFOBT [8]. The International Agency for Research on Cancer also concludes with sufficient evidence that CRC mortality reduction can be achieved through stool-based tests (gFOBT and FIT) and endoscopy (FS and colonoscopy) and the benefits outweigh the harms due to screening [9].

This study investigates the CRC burden in 28 European Union (EU) countries from 1990 to 2019, focusing on age-standardized rates, male–female differences and age-specific patterns of incidence and mortality. The data on CRC burden for 28 EU countries are retrieved from the Global Burden of Disease (GBD) 2019 study [10–12]. Apart from focusing on absolute counts and age-standardized rates of incidence and mortality, we also examine the changing patterns of disability-adjusted life years (DALYs) and its age-standardized rate. The changing trends of CRC burden are examined using joinpoint regression analysis. Previous studies examined the CRC burden at the global level [13] or different regions (Asia [14]; Central and South America [15]; Sub-Saharan Africa [16]); however, as per the best of our knowledge, this is the first study providing detailed epidemiological analysis of temporal patterns of CRC burden for all EU countries in the last 3 decades. This comparable and comprehensive examination of the CRC burden and its temporal patterns in EU countries might be important for readers, including medical doctors, researchers, and policymakers.

Data and methods

The data pertaining to incidence, mortality, and DALYs were procured from the Global Burden of Disease 2019 study from 1990 to 2019 (http://ghdx.healthdata.org/gbd-results-tool). We included all EU countries under analysis (including United Kingdom). The age-standardized incidence rate (ASIR), the age-standardized mortality rate (ASMR), and age-standardized DALYs rate (ASDALR) were expressed per 100,000 person-years, and were calculated using GBD world standard age population structure [11]. To understand the CRC burden by age, we aggregated the CRC burden in three broad age groups: 15–49 years, 50–70 years (middle-aged), and 70-plus (elderly). The rationale for categorizing the CRC burden in these three age groups is to understand how the CRC burden varies for adolescents and young adults (AYA), screening age-adults (50–70 years), and older age groups.

The GBD estimation framework is described in GBD publications [10–12]. Here, we briefly describe how GBD 2019 arrived at different estimates. The GBD estimation begins with collection of all data from sources, such as vital registration, sample registration system, and cancer registries. The data from vital registration and cancer registries are used in conjunction with the cause of death ensemble model to generate the mortality estimates; these mortality estimates are divided by mortality-to-incidence ratios to generate the estimates of incidence. Years of life lost (YLL) due to CRC are calculated by age-specific mortality estimates and standard life tables and years of life lived with disability (YLD) are calculated by multiplying disability weights with sequela-specific estimates. The DALYs are calculated as the sum total of YLL and YLD, and one DALY is interpreted as 1 year of healthy life lost.

Joinpoint analysis

Joinpoint regression analysis was performed to examine where the trends have changed significantly for age-standardized and age-specific rates. To calculate annual percent change, a regression line is fitted using the following equation:

$$\ln(S) = a + by.$$

In the above equation, *ln* denotes natural logarithm, S represents the metric for which APC is to be calculated, and y denotes the calendar year. The APC is calculated as $100*(e^b - 1)$. The average annual percent change (AAPC) is calculated as the weighted average of APCs, with weights being equal to the length of each segment over the specified fixed interval. The formula for calculation of AAPC is

$$\text{AAPC} = 100 \times \left\{ \exp\left(\frac{\sum w_i b_i}{\sum w_i} - 1 \right) \right\}.$$

In the above equation, w_i denotes the weight of each joinpoint, which is equal to length of the ith joinpoint segment, and b_i denotes the slope coefficient of ith joinpoint segment.

Results

Aggregate burden of CRC in European Union

Number of new cases of CRC in EU, for both sexes combined, increased by 70.2% from 261,306 in 1990 to 444,872

in 2019, whereas percent change in number of deaths was smaller (36.8%) between 1990 (155,823) and 2019 (213,174) (Fig. 1A). CRC DALYs decreased by 20.8% from 3.8 million in 1990 to 3.2 million in 2019. The ASIR increased by 11.9% from 37.8/100,000 to 42.3/100,000 between 1990 and 2019; in contrast, ASMR decreased by 16.9% (1990: 22.4/100,000; 2019: 18.6/100,000) and ASDALR decreased by 18.6% (1990: 472.9/100,000; 2019: 385.1/100,000) in the study period (Fig. 1B).

Sex-wise, the number of incident cases were similar between the two sexes in 1990 (males: 130,796; females: 130,510); however, the gap between males and females increased between 1990 and 2019, such that by 2019, males (248,954) had higher number of cases than females (195,918). Similarly, CRC deaths in females (80,192) were higher in 1990 than males (75,631); by 2019, however, CRC deaths in males (114,364) were higher than females (98,810). Notably, the age-standardized rates of CRC were consistently higher in males than females in all years between 1990 and 2019, with age-standardized rates tapering post-2005 period (Fig. 1B).

Country-wise CRC burden in 2019

Germany (78,951), Italy (60,514), and the United Kingdom (54,429) were the top three countries in terms of all-age incident cases in 2019 (Table 1). In terms of deaths, Germany (37,552) was again the leading country, with France (25,497) recording the second-highest deaths in 2019, followed by Italy (24,877). Germany (647,921), Italy (436,750), and United Kingdom (433,585) were again the leading countries in terms of DALYs. In 2019, 11/28 countries had higher ASIR than the EU as a whole, led by Slovakia (56.4/100,000) and Netherlands (55.5/100,000) (Table 1). Hungary was the leading country in terms of age-standardized mortality rate (ASMR: 28.6/100,000) and age-standardized DALYs rate (ASDALR: 630.3/100,000). Lithuania (29.2/100,000) had the lowest ASIR, whereas Finland had the lowest ASMR (12.3/100,000) and ASDALR (253.6/100,000) in 2019.

Male age-standardized rates were generally higher than females for almost all countries for all metrics with the male–female ratio varying between 1.5 and 2.0 in most countries (Fig. 2). In terms of gender-wise differences in age-standardized rates, Hungary was again the leading country in both males and females in terms of ASMR (male: 40.9/100,000; female: 20.5/100,000) and ASDALR (male: 884.5/100,000; female: 443.2/100,000). Slovakia (77.5/100,000) had the highest ASIR in males, whereas The Netherlands (47.6/100,000) had the highest ASIR among females in 2019. Finland and Latvia had male ASIR of CRC at 37.5/100,000, almost half that of Slovakia. Similarly, male ASMR was the lowest in Finland (14.9/100,000) and nearly 2.5 times lower than Hungary. Similar country-level differences in age-standardized rates were also visible among females (Fig. 2).

Temporal patterns by sex and country, 1990–2019

The percent change in age-standardized incidence and mortality rate in the EU and 28 countries between 1990 and 2019 are presented in Fig. 3 (the temporal patterns of ASIR and ASMR are shown in Supplementary Figs. S1–S2). Between 1990 and 2019, Romania registered the largest increases in ASIR (males: 138.4% vs females: 79.2%), ASMR (males: 82.3vs females: 31.9%), and ASDALR (males: 66.8% vs females: 22.8%), and Austria registered the largest reduction in both males (ASIR: −38.3%; ASMR: −53.5%; ASDALR: −55.4%) and females (ASIR: −33.6%; ASMR: −50.1%; ASDALR: −53.5%). Interestingly, positive percent changes in age-standardized rates were higher in males than females and negative percent changes were smaller in males than females, such that gap between male–female age-standardized rates widened in most countries between 1990 and 2019 (Supplementary Figs. S1–S2). Notably, while ASIR increased for both sexes in majority of countries, the ASMR and ASDALR registered decreases in most countries, even in those countries (e.g., Denmark, Finland, and Sweden) where ASIR increased between 1990 and 2019. Seven out of 28 countries (Croatia, Greece, Spain, Portugal, Lithuania, Hungary, and Slovenia) registered increases in ASMR and ASDALR in males and registered reductions in females between 1990 and 2019.

In total, 25/28 countries registered reductions in ASMR in females, while 15/28 countries experienced reductions in ASMR in males between 1990 and 2019 (Fig. 3).

Age-specific distribution of CRC burden in EU

The temporal patterns of absolute counts and age group specific rates (per 100,000) in three age groups are provided in Supplementary Figs. S3-5. In 15–49 year age group, incident cases and incident rate (per 100,000) increased till mid-2000s and stagnated thereafter; correspondingly, number of death counts and death rates stagnated before mid-2000s and decreased thereafter in both sexes (Supplementary Fig. S3). The age-specific rates increased or stagnated between 1990 and mid-2000s and decreased thereafter in 15–49 age group. The death rate as well as DALYs' rate in the age group of 15–49 years increased in males and decreased in females throughout the study period. In 50–69 age group, number of cases increased continuously between 1990 and 2019, whereas number of deaths and DALYs increased slightly in males and decreased slightly in females (Supplementary Fig. S4). Among those above the age of 70 years, absolute count of cases, deaths, and DALYs

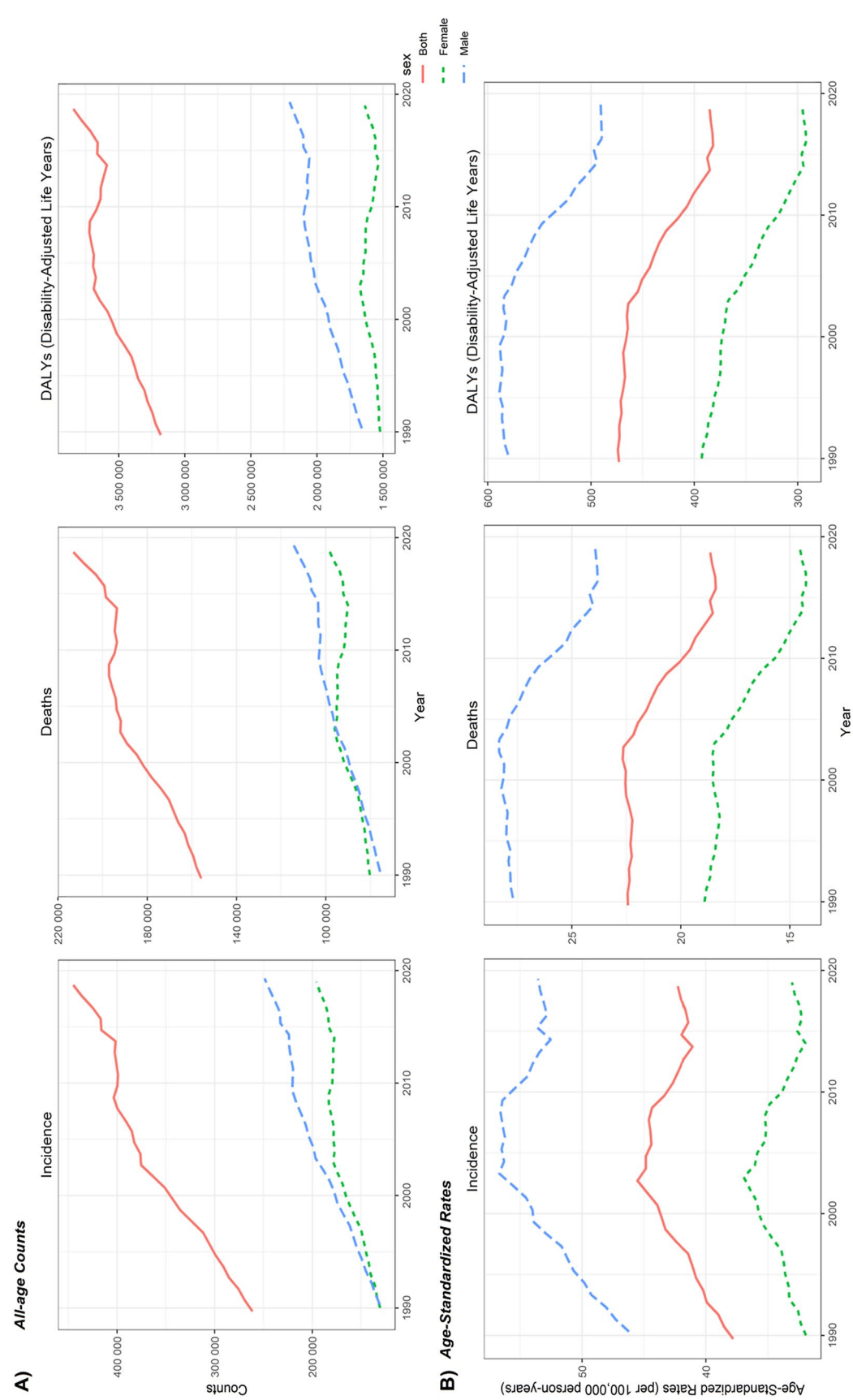


Fig. 1 Temporal patterns of colorectal cancer burden in European Union, 1990–2019. **A** All-age counts. **B** Age-standardized rates. Incidence: Incident Cases; Deaths: Number of Deaths; DALYs: Disability-adjusted Life Years. DALYs are calculated as the sum of years of life lost due to colorectal cancer and years of life lived with disability. Data Source: Global Burden of Disease 2019 Study

Table 1 Country-wise colorectal cancer burden in 2019

Country/region	Incidence	Deaths	DALYs	ASDALR	ASMR	ASIR
Austria	5847	2603	46,854	275.4	13.5	33.1
Belgium	8994	4325	73,674	337.5	17.1	39.3
Bulgaria	6199	3672	78,043	582.3	25.1	43.8
Croatia	4409	2308	44,362	529.2	25.3	50.5
Cyprus	739	265	5205	273.6	14.3	38.1
Czech Republic	8839	4678	91,540	452.1	21.6	42.1
Denmark	5577	2647	47,755	431.1	21.7	48.4
Estonia	1118	528	9665	384.2	18.5	42.2
Finland	3799	1616	28,876	253.6	12.3	31.5
France	52,277	25,497	422,215	328.7	16.4	38.5
Germany	78,951	37,552	647,921	360.7	18.0	41.4
Greece	7903	4000	66,397	297.9	14.9	33.2
Hungary	9873	5646	115,101	630.3	28.6	52.2
Ireland	3431	1377	26,562	362.7	18.0	46.0
Italy	60,514	24,877	436,750	331.0	15.8	43.5
Latvia	1214	746	14,162	377.4	17.6	30.5
Lithuania	1669	1024	19,673	365.6	16.9	29.2
Luxembourg	372	170	3075	313.5	16.0	37.2
Malta	303	136	2547	289.7	14.1	32.9
Netherlands	18,774	7774	142,372	435.3	21.8	55.4
Poland	24,277	17,768	350,819	518.0	24.7	35.0
Portugal	10,245	5189	92,569	421.3	20.1	45.3
Romania	13,039	7577	163,169	474.4	20.2	36.2
Slovakia	5203	2433	51,875	571.6	26.3	56.4
Slovenia	1726	913	16,809	408.2	19.8	41.1
Spain	47,432	20,011	350,229	387.3	18.8	50.1
Sweden	7720	3623	63,293	322.7	16.0	37.0
United Kingdom	54,429	24,220	433,585	365.2	18.1	43.6
European Union	444,872	213,174	3,845,097	385.2	18.7	42.3

Incidence: All-age incident cases, *Deaths* All-age deaths, *DALYs* Disability-adjusted life years, *ASIR* Age-standardized Incidence Rate, *ASMR* Age-standardized Mortality Rate, *ASDALR* Age-standardized DALYs' rate. The age-standardized rates are expressed per 100,000 population and the data are compiled from the Global Burden of Disease 2019 study

mainly increased between 1990 and 2019, whereas the age-specific rates of incidence, deaths, and DALYs increased or stagnated before 2000 and declined thereafter in the 70-plus age group. Among those aged 50–69 years, the incidence rate stagnated after mid-1990s whereas age-specific mortality and DALYs rate declined; in contrast, the rates increased or stagnated till 2003 among those aged 70 years and above and stagnated thereafter. Interestingly, there is a resurgence of incidence rate in the post-2014 period among adolescents and young adults (aged 15–49 years) and older adults (aged 70 years and above) (Supplementary Figs. S3–5).

Joinpoint analysis

Table 2 presents the joinpoint regression results for age-standardized rates as well as age-specific rates (15–49, 50–69, and 70-plus). For full period (1990–2019), the AAPC was positive for ASIR (0.4% [95% confidence interval 0.2%–0.5%]), whereas ASMR and ASDALR registered negative AAPC of −0.6% [−0.8% to −0.5%] and −0.7% [−0.8% to −0.7%], respectively. The breakpoints as per joinpoint regression yielded similar results for ASIR, ASMR, and ASDALR. Notably, ASIR (1.4% [1.3%–1.4%]) increased the most in 1990–2003, whereas ASMR increased by only 0.1% [0.0–0.2%] per year and ASDALR registered a decline of −0.2% [−0.2 to −0.1%] per year during this period. All three metrics registered a decline during 2003–14, whereas only ASIR has been increasing significantly (0.5 [0.1–0.8]) in the most recent period (2014–19).

For full period (1990–2019), the incident rate registered positive AAPC in all three age groups with the fastest rise occurring in 15–49 years; the mortality and DALYs' rate registered negative AAPC in all three age groups (Table 2, panel B–D). Among those aged 15–49 years, all three

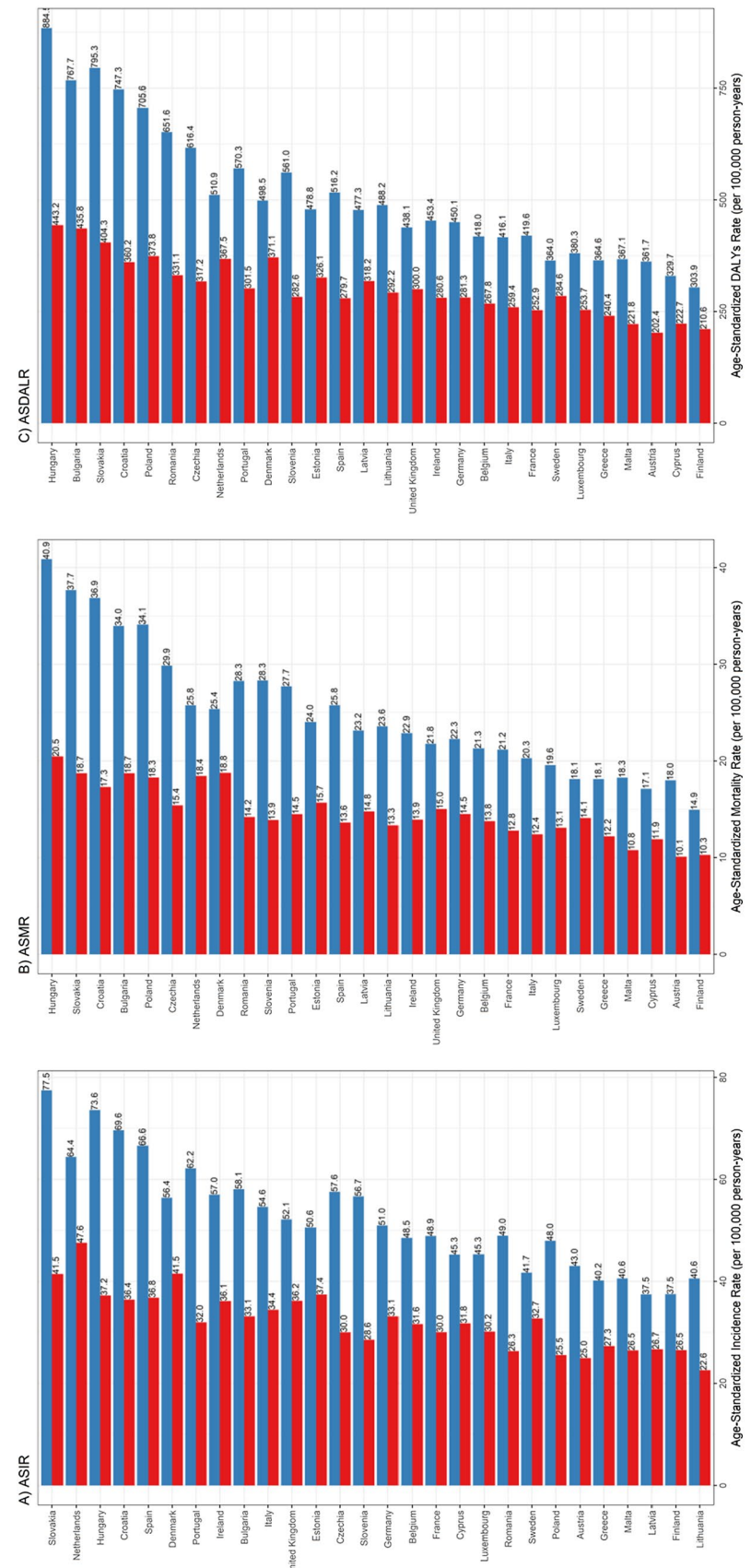


Fig. 2 Country-wise age-standardized rates of colorectal cancer in 2019: **A** ASIR, **B** ASMR, and **C** ASDALR. *ASIR* Age-standardized Incidence Rate, *ASMR* Age-standardized Mortality Rate, *ASDALR* Age-standardized DALYs rate. The age-standardized rates are expressed per 100,000 population and the data is compiled from the Global Burden of Disease 2019 study

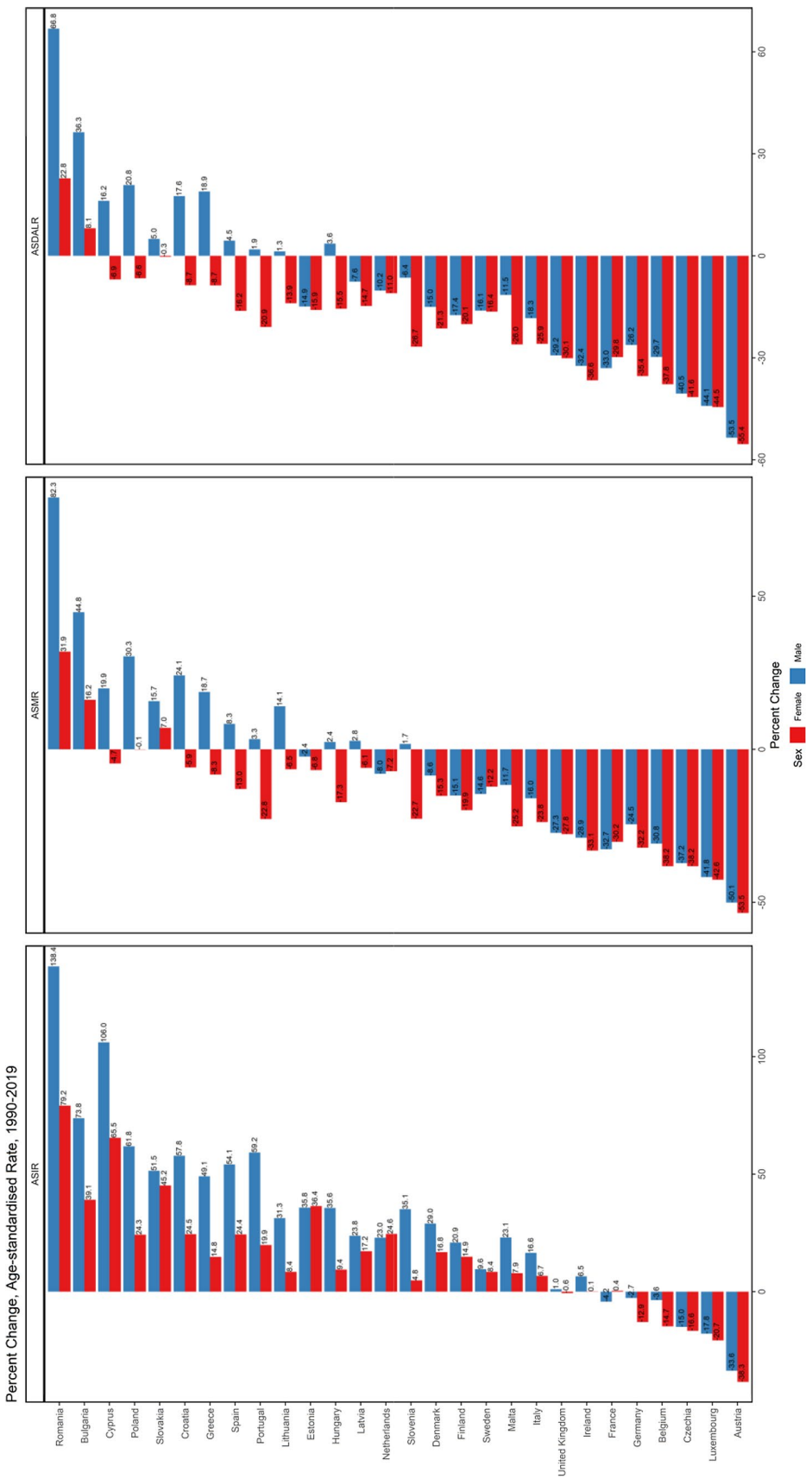


Fig. 3 Country and sexwise percent change in colorectal cancer age-standardized rates between 1990 and 2019. *ASIR* Age-standardized Incidence Rate, *ASMR* Age-standardized Mortality Rate, *ASDALR* Age-standardized DALYs rate. The age-standardized rates are expressed per 100,000 population and the data are compiled from the Global Burden of Disease 2019 study

Table 2 Joinpoint regression results of colorectal cancer

Variable	Trend 1		Trend 2		Trend 3		Trend 4	AAPC	
	Year (start –end)	APC	Year (start –end)	APC	Year (start –end)	APC	Year (start –end)	APC	AAPC
Panel A. Age-standardized rate (per 100,000)									
ASIR	1990–2003	1.4* (1.3–1.4)	2003–09	−0.4* (−0.8 to −0.1)	2009–14	−1.3* (−1.8 to −0.9)	2014–19	0.5* (0.1–0.8)	0.4* (0.2–0.5)
ASMR	1990–2003	0.1* (0.0–0.2)	2003–08	−1.4* (−1.9 to −0.9)	2008–14	−2.1* (−2.5 to −1.8)	2014–19	0.1 (−0.3 to 0.4)	−0.6* (−0.8 to −0.5)
ASDALR	1990–2003	−0.2* (−0.2 to −0.1)	2003–09	−1.4* (−1.6 to −1.2)	2009–14	−2.0* (−2.3 to −1.7)	2014–19	−0.1 (−0.3 to 0.2)	−0.7* (−0.8 to −0.7)
Panel B. Age-specific rate in 15–49 years									
Incidence rate	1990–2003	2.1* (2.1–2.2)	2003–09	0.3* (−0.0 to 0.7)	2009–14	−1.4* (−1.9 to −0.9)	2014–19	0.8* (0.4–1.2)	0.9* (0.8–1.1)
Mortality rate	1990–2002	0.4*(0.3–0.4)	2002–08	−1.0* (−1.3 to −0.7)	2008–14	−2.3* (−2.6 to −1.9)	2014–19	0.3 (0.0 to 0.7)	−0.5* (−0.6 to −0.4)
DALYs' rate	1990–2002	0.4* (0.2–0.4)	2002–08	−0.9* (−1.2 to −0.6)	2008–14	−2.3* (−2.6 to −1.9)	2014–19	0.4* (0.1–0.7)	−0.5* (−0.6 to −0.3)
Panel C. Age-specific Rate in 50–69 years									
Incidence rate	1990–1995	2.0* (1.5–2.4)	1995–2009	0.2* (0.1–0.3)	2009–13	−0.5 (−1.6 to 0.5)	2013–19	0.2 (−0.2 to 1.0)	0.4* (0.2–0.6)
Mortality rate	1990–1996	0.1 (−0.1 to 0.2)	1996–2001	−1.7* (−2.0 to -1.4)	2001–14	−1.1* (−1.2 to −1.1)	2014–19	−0.2* (−0.5 to -0.0)	−0.8* (−0.9 to −0.8)
DALYs' rate	1990–1996	0.1 (−0.1 to 0.2)	1996–2000	−1.5* (−2.0 to −1.0)	2000–14	−1.1*(−1.1 to −1.0)	2014–19	−0.5* (−0.7 to −0.2)	−0.8* (−0.9 to −0.7)
Panel D. Age-specific rate in 70-plus years									
Incidence rate	1990–1997	1.2* (0.8–1.5)	1997–2003	2.3* (1.7–3.0)	2003–14	−1.1* (−1.4 to −0.9)	2014–19	0.6* (0.0–1.3)	0.4* (0.3–0.6)
Mortality rate	1990–1996	−0.3 (−0.6–0.1)	1996–2003	1.2*(0.9–1.6)	2003–14	−1.5* (−1.7 to −1.4)	2014–19	0.6* (0.0–1.0)	−0.3* (−0.4 to −0.1)
DALYs' rate	1990–1996	−0.1 (−0.5 to 0.2)	1996–2003	1.0* (0.6 to 1.3)	2003–14	−2.0* (−2.1 to −1.8)	2014–19	0.3 (−0.2 to 0.7)	−0.5* (−0.6 to −0.4)

ASIR Age-standardized Incidence Rate, *ASMR* Age-standardized Mortality Rate, *ASDALR* Age-standardized DALYs rate. *AAPC* Average Annual Percent Change. Incidence, Mortality and DALYs rate in three broad age groups are expressed per 100,000 population. Data Source: Global Burden of Disease 2019 Study

metrics registered increases in the 1990–2002/03 period, whereas these three metrics registered reductions in the 2008–14 period. In the most recent period (2014–19), all three metrics (incidence, mortality, and DALY rate) have been increasing in the 15–49 age group. In the 50–69 age group, the incident rate increased the fastest (2.0% [1.5–2.4]) in 1990–95 period, whereas mortality rate decreased by −1.7% [−2.0 to −1.4%] and −1.1% [−1.2 to −1.1%], respectively, during 1996–2001 and 2001–14. Notably, mortality and DALYs' rate decreased significantly from 1996 to 2019 in the 50–69 age group. In 70-plus age group, the incidence rate increased in all periods except 2003–14 with an overall AAPC of 0.4% [0.3–0.6] during 1990–2019. During 1996/97–2003 period, all three metrics (age-specific incidence, mortality, and DALY rate) registered increases in the 70-plus

age group, whereas these three metrics registered decrease in 2003–14 period in this age group.

Discussion

The absolute burden of CRC increased in EU and as well as in individual countries, whereas apart from the incidence rate, other age-standardized rates (mortality and DALYs) decreased in the last three decades. The reductions in mortality rates are consistent with similar findings from a previous study in Europe [17]. As per joinpoint analysis, from 1990 to mid-2000s and 2014–19, the ASIR increased in EU, whereas ASMR and ASDALR registered statistically significant decreases in the 2003–14 period. Notably, in 2014–19 period, the age-specific rates have been increasing in the 15–49 age-group.

Between 1990 and 2019, ASIR increased in 18/28 countries, led by Romania, Bulgaria, and Cyprus, whereas ASMR and ASDALR reduced for both sexes in the majority of EU countries, with a few countries (e.g., Romania and Bulgaria) also registered increases in ASMR and ASDALR for both males and females. For several countries (e.g., Germany, Austria, Czechia, and Denmark), the ASMR observed a turning point in mid-2000s; this trend reversal can be attributed to population-based screening initiated in these countries. For instance, CRC incidence with distant metastasis has decreased in Germany due to colonoscopy screening beginning in 2002 [18]. A recent study has also attributed the divergent trends in CRC incidence and mortality in Europe to different years in which CRC screening was initiated in different countries [19].

The highest incidence and mortality rates were observed in Hungary and Slovakia, where CRC incidence rates have been growing recently due to changing lifestyles [20]. More than 95% of CRC cases can be treated if diagnosed early or premalignant polyps are detected and removed [21]; in this endeavour, and mass-screening programs have substantially reduced CRC mortality through early detection [22, 23]. Notably, the age-standardized rates of CRC decreased the most in countries such as Austria, which have long-standing screening programs [19]. Late-disease detection entails poor prognosis and results in higher treatment costs, with late-stage disease costing tenfold higher than early-stage disease [24]. As per a white paper on digestive cancers in Europe, if EU countries could increase CRC detection in stage 1 from 13% to 50%, 130,000 more lives and €3 billion would be saved on treatment costs [24].

Between 1990 and 2019, the incidence rate of CRC grew the fastest in 15–49 age group (AAPC: 0.9%), followed by the 50–69 and 70-plus age groups, both of these age groups registered AAPC of 0.4% per year. A few previous studies have also highlighted the rising incidence of CRC in young adults [25, 26]. CRC among adolescents can be linked to genetic or familial risk, whereas the reasons for rising CRC among young adults are hard to pinpoint, researchers have attributed the increase to behavioral risk factors including obesity, diet, early smoking initiation, and lack of physical activity [26]. Rising obesity rates in young adults can be one of the possible reasons for rising CRC incidence as weight gain is implicated as an important risk factor toward rising CRC [27]. Although smoking rates have decreased worldwide [28], yet smoking initiation is still unacceptably high in adolescents in Europe and is increasing in those aged 15 years or below [29].

In European Prospective Investigation into Cancer and Nutrition (EPIC) cohort study, it was found that healthy lifestyle factors are linked with lower CRC incidence [30]. Previous studies on the association between CRC risk and Mediterranean diet, characterized by frequent consumption of fruits, vegetables, legumes, olive oil, fish, nuts and cereals and modest amounts of red wine, dairy products and meat [31], have found that Mediterranean diet is associated with lower risk of CRC incidence [32–36]. Previous studies have also found that a proinflammatory diet is associated with increased CRC risk [37–41], which suggests the role of diet in augmenting the CRC risk.

The treatment choice of CRC depends upon several factors such as stage and size of tumor, its location, and patients' characteristics such as age [42]. According to the European Society for Medical Oncology (ESMO), descriptions concerning CRC are divided into four categories, namely, early colon cancer, rectal cancer, metastatic colorectal cancer, and familial risk colorectal cancer [43]. The treatment regimens encompass surgery to remove the tumor and any affected lymph nodes for most patients with localized CRC (early and locally advanced). Chemotherapy is usually offered post-operatively to patients with localized disease with lymph-node involvement to reduce the risk of recurrence. Patients with metastatic disease must receive systemic treatment, including targeted therapies, depending on the disease characteristics (including molecular biology). In selected cases, treatment for metastatic disease can include local treatment (surgery, radiotherapy, and interventional radiology) and can aim to be curative [44]. According to the ESMO guidelines, patients with stage III CRC are recommended to receive adjuvant chemotherapy with 5-FU and OX (infusional fluorouracil + l-leucovorin + oxaliplatin (FOLFOX) or capecitabine + oxaliplatin (CapeOX)) as well as capecitabine (Cape) or fluorouracil + l-leucovorin (5-FU + l-LV). However, given that side effects such as diarrhea or peripheral neurotoxicity are often reported with OX, these combination regimens are not recommended to all patients with stage III CRC. In the ESMO guidelines, induction timing of adjuvant chemotherapy is described as "as early

as possible," starting from the third week up to a maximum of 8-12 weeks after surgery [43].

Surgical resection is the mainstay of definitive treatment of CRC and a more aggressive surgical approach of treating elderly patients can improve treatment outcomes and survival rates in Europe [45]. Chemotherapy for CRC is also suggested in an adjuvant manner or in cases where resection is not feasible. A study on 1.3 million cases from 22 cancer registries in Europe found that surgery alone or in combination with radiotherapy or systemic therapy is the mainstay of treatment of CRC [46]. A study comparing CRC treatment in Europe, Australia, and New Zealand found a decreasing trend toward abdominoperineal (APR) resection and an increasing trend toward sphincter sparing surgery such as mesorectal excision and lower anterior resection [47]. Mesorectal excision has not only improved the survival rates from rectal cancer [48, 49], but sphincter preservation surgery has also been found to be superior to APR in terms of quality of life of the CRC patient [50]. An improvement in surgical techniques for localized tumors and new treatment protocols (including adjuvant therapy) may have also helped in reducing CRC mortality [51, 52].

In the European Union as well as individual countries, males had higher age-standardized rates of CRC than females. We have also observed that male–female differences have grown wider between 1990 and 2019, with age-standardized rates and absolute counts being smaller in females than males. A previous study has also highlighted the greater burden of CRC in males than females [53]. The male–female differences are not specific to EU only, but such differences also exist at the global level and in other regions [54]. A population-based study in Germany has also found survival advantage of CRC among females compared to males, mostly in young and middle-aged patients [55]; in United Kingdom, however, the gender-wise differences are believed to be due to incidence differences arising out of factors affecting the incidence risk of CRC [56].

A previous study comparing 5-year survival rates between the United States and Europe has found lower 5-year survival rates in Eastern Europe than the United States and northern and Western Europe [45]. The lower survival rates in Eastern Europe than the United States can be attributed to early stage detection and extensive usage of surgery and adjuvant treatment in the United States [45]. A recent study has suggested that lower survival rates in Eastern Europe can be attributed to the lower percentage of cancers detected in the early stages than in the metastatic stages and fewer resources allocated to health in Eastern Europe than other European regions [57]. Therefore, increased resource allocation and early detection through screening are the cornerstones of improving CRC outcomes in central and Eastern Europe [58]. Among European countries, the mortality rates have reduced and early stage detection increased in those countries which have long-standing screening programs [19].

CRC screening is one of the most critical determinants of early stage detection and hence survival rates. By 2015, 24/28 EU countries had either established or were preparing a nationwide organized or opportunistic screening program for CRC [59]. A few countries offer population-wide screening using FS or colonoscopy in Europe, or a few have started offering opportunistic or pilot-based colonoscopy screening [58]. A systematic review of the efficacy of screening in Europe found that screening using gFOBT is associated with 8–16% reduction in CRC mortality and FS was associated with 21–30% reduction in CRC mortality [58]. A study [19] reported varying percentages of the eligible population who have undergone either a fecal test within 2 years or colonoscopy within ten years across European countries; Germany (70.9%), Austria (68.8%), UK (66.7%), Slovenia (65%), Denmark (47%), Ireland (42.7%), Finland (34.8%), Belgium (32.7%), Lithuania (27.7%), and The Netherlands (24.9%). FIT screening reduced the CRC incidence and mortality risk in Italy [60]. In Germany, too, screening through colonoscopy has been shown to have great potential for preventing and early detection of CRC, with lower risk of over diagnosis [61]. A few studies have also highlighted that FS screening can reduce the incidence rates as compared to gFOBT and the results were not statistically significant in case of gFOBT [62–64]. For countries with more recent screening implementation (such as Belgium, Denmark, Ireland, Lithuania, Netherlands, Slovenia), an increase in the proportion of stage I CRC was observed after implementing the organized FIT-based programs. Although screening using different methods is more cost-effective than no screening [65], among different screening modalities, there is a disagreement about the most cost-effective method or the one with best incremental cost-effectiveness ratio [66].

The mere presence of screening might not ensure improved CRC survival rates until high screening uptake and participation levels are ensured. A European council report suggests 65% participation rate in population-level screening as desirable in a country, yet only 14% of EU citizens aged 50–74 years have the opportunity to participate in the formal population-based screening program [24]. Netherlands and Slovenia ensured high participation rates in the screening programs, resulting in higher percentage of patients getting diagnosed in stage I [24, 67]. Slovenia set up the national screening program for CRC in 2009, targeting the population aged 50–69 years (the age group was later broadened to include those aged 69–74 years), which has been instrumental in the early detection (stage 1/2) of CRC cases and thereby lowering cancer stages and incidence [67].

Conclusion

CRC incidence rates have been increasing in EU countries in the last 3 decades and mortality rates, although decreasing, are still unacceptably high. As several CRC deaths can be prevented through early detection, the focus must be on early stage detection through high-quality screening with higher participation rates. Apart from screening, the CRC burden can be reduced through preventive approaches based on modifiable risk factors. There must be educational campaigns focusing on benefits of a healthy diet, physical activity, awareness of disease symptoms, and benefits of early detection. In recent years, CRC incidence has been increasing among those aged less than 50 years, which demands more research to understand the reasons behind the rising CRC incidence in young adults and to pinpoint the risk factors. Moreover, there must be more awareness of these trends of rising CRC incidence among clinicians and general public to curtail this rising trend.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10147-022-02182-0.

References

1. Sung H, Ferlay J, Siegel RL et al (2021) Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin 71(3):209–249
2. Johnson CM, Wei C, Ensor JE et al (2013) Meta-analyses of colorectal cancer risk factors. Cancer Causes Control 24(6):1207–1222
3. World Cancer Research Fund/American Institute for Cancer Research (2018) Food, nutrition, physical activity, and Cancer: a global perspective. Continuous Update Project Expert Report
4. Vasen HF, Möslein G, Alonso A et al (2010) Recommendations to improve identification of hereditary and familial colorectal cancer in Europe. Fam Cancer 9(2):109–115
5. Hampel H, Frankel WL, Martin E et al (2005) Screening for the Lynch syndrome (hereditary nonpolyposis colorectal cancer). N Engl J Med 352(18):1851–1860
6. Henderson RH, French D, Maughan T et al (2021) The economic burden of colorectal cancer across Europe: a population-based cost-of-illness study. Lancet Gastroenterol Hepatol 6(9):709–722
7. Council of the European Union. Council recommendation of 2 December 2003 on cancer screening (2003/878/EC) (2003) Off J Eur Union L 327:34e8
8. European Colorectal Cancer Screening Guidelines Working Group (2013) European guidelines for quality assurance in colorectal cancer screening and diagnosis: overview and introduction to the full supplement publication. Endoscopy 45(01):51–59
9. Lauby-Secretan B, Vilahur N, Bianchini F et al (2018) The IARC perspective on colorectal cancer screening. N Engl J Med 378(18):1734–1740
10. Vos T, Lim SS, Abbafati C et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396:1204–1222
11. Wang H, Abbas KM, Abbasifard M et al (2020) Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the global burden of disease study 2019. Lancet 396:1160–1203
12. Murray CJ, Aravkin AY, Zheng P et al (2020) Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the Global Burden of Disease Study 2019. Lancet 396(10258):1223–1249
13. Safiri S, Sepanlou SG, Ikuta KS et al (2019) The global, regional, and national burden of colorectal cancer and its attributable risk factors in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. Lancet Gastroenterol Hepatol 4(12):913–933
14. Pourhoseingholi MA (2012) Increased burden of colorectal cancer in Asia. World J Gastrointest Oncol 4(4):68
15. Sierra MS, Forman D (2016) Burden of colorectal cancer in Central and South America. Cancer epidemiol 44:S74-81
16. Katsidzira L, Gangaidzo I, Thomson S et al (2017) The shifting epidemiology of colorectal cancer in sub-Saharan Africa. Lancet Gastroenterol Hepatol 2(5):377–383
17. Ouakrim DA, Pizot C, Boniol M et al (2015) Trends in colorectal cancer mortality in Europe: retrospective analysis of the WHO mortality database. BMJ. https://doi.org/10.1136/bmj.h4970
18. Oppelt KA, Luttmann S, Kraywinkel K et al (2019) Incidence of advanced colorectal cancer in Germany: comparing claims data and cancer registry data. BMC Med Res Methodol 19:142
19. Cardoso R, Guo F, Heisser T et al (2021) Colorectal cancer incidence, mortality, and stage distribution in European countries in the colorectal cancer screening era: an international population-based study. Lancet Oncol 22(7):1002–1013
20. Arnold M, Karim-Kos HE, Coebergh JW et al (2015) Recent trends in incidence of five common cancers in 26 European countries since 1988: analysis of the European Cancer Observatory. Eur J Cancer 51:1164e87
21. Pawa N, Arulampalam T, Norton J (2011) Screening for colorectal cancer: established and emerging modalities. Nat Rev Gastroenterol Hepatol 8:711–722
22. Burt RW (2010) Colorectal cancer screening. Curr Opin Gastroenterol 26:466–470
23. Gupta AK, Brenner DE, Turgeon DK (2008) Early detection of colon cancer: new tests on the horizon. Mol Diagn Ther 12:77–85
24. Digestive Cancers Europe (2019) White Paper Colorectal Cancer Screening in Europe. Available at https://www.digestivecancers.eu/wp-content/uploads/2020/02/466-Document-DiCEWhitePaper2019.pdf (Accessed 3.12.2021)
25. Siegel RL, Torre LA, Soerjomataram I et al (2019) Global patterns and trends in colorectal cancer incidence in young adults. Gut 68(12):2179–2185
26. Vuik FE, Nieuwenburg SA, Bardou M et al (2019) Increasing incidence of colorectal cancer in young adults in Europe over the last 25 years. Gut 68(10):1820–1826
27. Karahalios A, English DR, Simpson JA (2015) Weight change and risk of colorectal cancer: a systematic review and meta-analysis. Am J Epidemiol 181:832–845
28. Reitsma MB, Flor LS, Mullany EC et al (2021) Spatial, temporal, and demographic patterns in prevalence of smoking tobacco use and initiation among young people in 204 countries and territories, 1990–2019. Lancet Public Health 6(7):e472–e481
29. Marcon A, Pesce G, Calciano L et al (2018) Trends in smoking initiation in Europe over 40 years: a retrospective cohort study. PLoS ONE 13(8):e0201881
30. Aleksandrova K, Pischon T, Jenab M et al (2014) Combined impact of healthy lifestyle factors on colorectal cancer: a large European cohort study. BMC Med 12:168
31. Trichopoulou A, Costacou T, Bamia C et al (2003) Adherence to a Mediterranean diet and survival in a Greek population. N Engl J Med 348(26):2599–2608

32. Farinetti A, Zurlo V, Manenti A et al (2017) Mediterranean diet and colorectal cancer: A systematic review. Nutrition 43:83–88
33. Jones P, Cade JE, Evans CE et al (2017) The Mediterranean diet and risk of colorectal cancer in the UK Women's Cohort Study. Int J Epidemiol 46(6):1786–1796
34. Rosato V, Guercio V, Bosetti C et al (2016) Mediterranean diet and colorectal cancer risk: a pooled analysis of three Italian case–control studies. Br J Cancer 115:862–865
35. Donovan MG, Selmin OI, Doetschman TC et al (2017) Mediterranean diet: prevention of colorectal cancer. Front nutr 4:59
36. Reedy J, Mitrou PN, Krebs-Smith SM et al (2008) Index-based dietary patterns and risk of colorectal cancer: the NIH-AARP Diet and Health Study. Am J Epidemiol 168(1):38–48
37. Tabung FK, Steck SE, Ma Y et al (2015) The association between dietary inflammatory index and risk of colorectal cancer among postmenopausal women: results from the Women's Health Initiative. Cancer Causes Control 26(3):399–408
38. Fan Y, Jin X, Man C et al (2017) Meta-analysis of the association between the inflammatory potential of diet and colorectal cancer risk. Oncotarget 8(35):59592
39. Zhang C, Wang W, Zhang D (2018) Association between dietary inflammation index and the risk of colorectal cancer: a meta-analysis. Nutr cancer 70(1):14–22
40. Shivappa N, Zucchetto A, Montella M et al (2015) Inflammatory potential of diet and risk of colorectal cancer: a case–control study from Italy. Br J Nutr 114(1):152–158
41. Shivappa N, Godos J, Hébert JR et al (2017) Dietary inflammatory index and colorectal cancer risk—a meta-analysis. Nutrients 9(9):1043
42. Marley AR, Nan H (2016) Epidemiology of colorectal cancer. Int J Mol Epidemiol Genet 7(3):105–114
43. Shinagawa T, Tanaka T, Nozawa H et al (2018) Comparison of the guidelines for colorectal cancer in Japan, the USA and Europe. Ann Gastroenterol Surg 2(1):6–12
44. Beets G, Sebag-Montefiore D, Andritsch E et al (2017) ECCO essential requirements for quality cancer care: colorectal cancer. A critical review. Crit Rev Oncol Hematol 110:81–93
45. Allemani C, Rachet B, Weir HK et al (2013) Colorectal cancer survival in the USA and Europe: a CONCORD high-resolution study. BMJ Open 3(9):e003055
46. Giusti F, Martos C, Neamtiu L et al (2020) 508P Evaluating colorectal cancer treatment patterns by stage in European cancer registries. Ann Oncol 31:S455
47. Chawla N, Butler EN, Lund J et al (2013) Patterns of colorectal cancer care in Europe, Australia, and New Zealand. J Natl Cancer Inst Monographs 2013(46):36–61
48. Kapiteijn E, Putter H, Van de Velde CJ (2002) Impact of the introduction and training of mesorectal excision on recurrence and survival of rectal cancer in the Netherlands. Br J Surg 89:1142–1149
49. Heald RJ (1995) Total mesorectal excision is optimal surgery for rectal cancer: a Scandinavian consensus. Br J Surg 82:1297–1299
50. Kang SB, Cho JR, Jeong SY et al (2021) Quality of life after sphincter preservation surgery or abdominoperineal resection for low rectal cancer (ASPIRE): A long-term prospective, multicentre, cohort study. Lancet Regional Health-Western Pacific 6:100087
51. Cunningham D, Atkin W, Lenz HJ et al (2010) Colorectal cancer. Lancet 375:1030–1047
52. Bosetti C, Levi F, Rosato V et al (2011) Recent trends in colorectal cancer mortality in Europe. Int J Cancer 129(1):180–191
53. Micheli A, Ciampichini R, Oberaigner W et al (2009) The advantage of women in cancer survival: an analysis of EUROCARE-4 data. Eur J Cancer 45(6):1017–1027
54. Sharma R (2020) An examination of colorectal cancer burden by socioeconomic status: evidence from GLOBOCAN 2018. EPMA J 11(1):95–117
55. Majek O, Gondos A, Jansen L et al (2013) Sex differences in colorectal cancer survival: population-based analysis of 164,996 colorectal cancer patients in Germany. PloS One 8(7):e68077
56. White A, Ironmonger L, Steele RJ et al (2018) A review of sex-related differences in colorectal cancer incidence, screening uptake, routes to diagnosis, cancer stage and survival in the UK. BMC Cancer 18(1):1–1
57. Minicozzi P, Walsh PM, Sánchez MJ et al (2018) Is low survival for cancer in Eastern Europe due principally to late stage at diagnosis? Eur J Cancer 93:127–137
58. Gini A, Jansen EE, Zielonke N et al (2020) Impact of colorectal cancer screening on cancer-specific mortality in Europe: A systematic review. Eur J Cancer 127:224–235
59. Schreuders EH, Ruco A, Rabeneck L et al (2015) Colorectal cancer screening: a global overview of existing programmes. Gut 64(10):1637–1649
60. Rossi PG, Vicentini M, Sacchettini C et al (2015) Impact of screening program on incidence of colorectal cancer: a cohort study in Italy. Am J Gastroenterol 110(9):1359–1366
61. Brenner H, Altenhofen L, Stock C et al (2015) Prevention, early detection, and overdiagnosis of colorectal cancer within 10 years of screening colonoscopy in Germany. Clin Gastroenterol Hepatol 13(4):717–723
62. Holme O, Loberg M, Kalager M, et al (2018) Long-term effectiveness of sigmoidoscopy screening on colorectal cancer incidence and mortality in women and men: a randomized trial. Ann Intern Med 168:775e82
63. Atkin W, Wooldrage K, Parkin DM et al (2017) Long term effects of once-only flexible sigmoidoscopy screening after 17 years of follow-up: the UK Flexible Sigmoidoscopy Screening randomised controlled trial. Lancet 389:1299e311
64. Segnan N, Armaroli P, Bonelli L et al (2011) Once-only sigmoidoscopy in colorectal cancer screening: follow-up findings of the Italian Randomized Controlled Trial—SCORE. J Natl Cancer Inst 103(17):1310–1322
65. Ran T, Cheng CY, Misselwitz B et al (2019) Cost-effectiveness of colorectal cancer screening strategies—a systematic review. Clin Gastroenterol Hepatol 17(10):1969–1981
66. Lansdorp-Vogelaar I, Knudsen AB, Brenner H (2011) Cost-effectiveness of colorectal cancer screening. Epidemiol rev 33(1):88–100
67. Novak Mlakar D, Kofol Bric T (2018) Slovenian national colorectal cancer screening–Programme SVIT. Eur J Public Health 28(suppl_4):cky213–786

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

Water Resources Management
https://doi.org/10.1007/s11269-022-03173-6

A Comparative Study of Data-driven Models for Groundwater Level Forecasting

R. Sarma[1] · **S. K. Singh**[1]

Received: 25 November 2021 / Accepted: 4 May 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
Irregular rainfall patterns and limited freshwater availability have driven humans to increase their dependence on groundwater resources. An essential aspect of effective water resources management is forecasting groundwater levels to ensure that sufficient quantities are available for future generations. Prediction models have been widely used to forecast groundwater levels at the regional scale. This study compares the accuracy of five commonly used data-driven models–Holt–Winters' Exponential Smoothing, Seasonal Autoregressive Integrated Moving Average, Multi-Layer Perceptron, Extreme Learning Machine, and Neural Network Autoregression for simulating the declining groundwater levels of three monitoring wells in the National Capital Territory of Delhi in India. The performance of the selected models was compared using coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Results indicate that Multi-Layer Perceptron had high R^2 while fitting the training data and least RMSE and MAE during testing, thus proving to be more accurate in forecasting than the other models. Multi-Layer Perceptron was used to forecast the groundwater level in the study wells for 2025. The results showed that the groundwater level will decline further if the current situation continues. Such studies help determine the appropriate model to be used for regions with limited available data. Additionally, predictions made for the future will help policymakers understand which areas need immediate attention in terms of groundwater management.

Keywords Groundwater level forecasting · Holt-winters' exponential smoothing · Seasonal ARIMA · Multi-layer perceptron · Extreme learning machine · Neural network autoregression

Highlights
- A number of different data-driven models exist for forecasting groundwater levels.
- In the present study, Multi-Layer Perceptron was the most accurate model for groundwater level forecasting.
- Accuracy measures - coefficient of determination, Root Mean Squared Error and Mean Absolute Error were compared to determine the most precise model.
- Multi-Layer Perceptron forecasts for 2025 showed a decline of 2-21 mbgl in the study wells.

✉ S. K. Singh
sksinghdce@gmail.com

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

1 Introduction

With rising dependence on groundwater to meet domestic, irrigation and industrial demands, trend analysis and simulation of water table behaviour have become important fields of study (Noori and Singh 2021). Researchers worldwide have developed and applied a number of different models to forecast groundwater levels, particularly for areas where recharge is low and extraction is high. Although models available for forecasting are abundant, selecting a suitable model that accurately simulates groundwater behaviour is complex (Takafuji et al. 2019). Several factors have to be taken into account for model choice and performance–objective of the study and data availability being primary (Salvadore et al. 2015; Barzegar et al. 2017).

Numerical modeling techniques have gained immense popularity in recent years, mainly due to their ability to forecast the likely impacts of water management solutions (Singh 2014; Sarma and Singh 2021a). Although largely popular, numerical models may be limited by large data requirements and complex computations (Aguilera et al. 2019). Univariate forecasting can be beneficial for data-scarce study areas. Zhang et al. (2018) present a comprehensive review of commonly used data-driven models for hydrological processes. They classified data-driven models as conventional, AI-based and hybrid models for streamflow prediction.

Conventional time series models such as Autoregressive Integrated Moving Average (ARIMA) and Seasonal Autoregressive Integrated Moving Average (SARIMA) are easy to use and widely popular (Mirzavand and Ghazavi 2015; Choubin and Malekian 2017; Rahaman et al. 2019;). The ARIMA model assumes that data at time t will directly correlate with previous data at t-1, t-2, …. and associated errors (Narayanan et al. 2013). Many studies around the world have compared the performance of ARIMA or SARIMA with other models such as Holt-Winters' Exponential Smoothing (HWES), Integrated Time Series model (ITS) and Artificial Neural Networks (ANN) (Shirmohammadi et al. 2013; Aguilera et al. 2019; Sakizadeh et al. 2019). The HWES model forecasting is based on Holt and Winters' basic structures (Holt 1957; Winters 1960) and can be applied on data with non-constant trends and seasonal variations (Yang et al. 2017). In the study by Yang et al. (2017) for a coastal aquifer in South China, HWES outperformed ITS and SARIMA.

In recent years, ANNs have seen widespread applications in forecasting hydrological parameters (Chen et al. 2020; Mozaffari et al. 2022). ANN replicates biological neuron processing and consists of an input layer, one or more hidden layers and an output layer. Lallahem et al. (2005) evaluated the feasibility of using ANNs for groundwater level predictions and concluded that ANNs, particularly MLPs with minimal lags and hidden nodes, gave the best simulation results. Yang et al. (2009) reported that Back-Propagation Artificial Neural Network outperformed ITS in simulating groundwater levels in China. Aguilera et al. (2019) compared the performance of the Prophet model with other forecasting techniques–seasonal naïve, linear model, exponential smoothing, ARIMA and neural network autoregression (NNAR) for groundwater level data in Spain. Recent years have seen the growing use of a novel method called Extreme Learning Machine (ELM) in hydrology (Kalteh 2019; Parisouj et al. 2020). In a study by Natarajan and Sudheer (2020), ELM had the best performance compared to ANN, GP and SVM for groundwater level prediction at six locations in Andhra Pradesh, India. Poursaeid et al. (2022) compared the performance of some mathematical and Artificial Intelligence (AI) models and concluded that the ELM method showed the best performance for groundwater level simulation.

In this study, two conventional models – Holt-Winters' Exponential Smoothing (HWES), Seasonal ARIMA and three AI-based models – Multi-Layer Perceptron (MLP), Extreme Learning Machine (ELM) and Neural Network Autoregression (NNAR) were applied on historical groundwater records of three monitoring wells in National Capital Territory of Delhi, India. The models were trained and tested as per standard procedure. The forecasting performance of each model was compared using accuracy measures–coefficient of determination (R^2), Root Mean Squared Error (RMSE) and Mean Absolute Error (MAE). Finally, the best model was used to predict groundwater depth in 2025. To the best of our knowledge, this comparative analysis has been done for the first time in the study area.

2 Materials and Methods

2.1 Data Driven Models

2.1.1 Holt-winters' Exponential Smoothing

The simple smoothing methodology given by Brown (1959) was limited by its inability to support a time series with trend components (Sakizadeh et al. 2019). Holt (1957) proposed the exponential smoothing method which incorporated a trend component given by the following equations:

Level

$$l_t = \alpha y_t + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{1}$$

Trend

$$b_t = \beta^* \left(l_t - l_{t-1}\right) + (1-\beta^*) b_{t-1} \tag{2}$$

Forecast

$$y_{t+h|t} = l_t + b_t h \tag{3}$$

where, l_t: level of time series at t time step, b_t:slope of the series, $y_{t+h|t}$: forecast for next h time steps, α, β^*: smoothing parameters (between 0 and 1). For a time series with variable seasonal fluctuations, simulations are made using the multiplicative form:

Level

$$l_t = \alpha \frac{y_t}{S_{t-m}} + (1-\alpha)(l_{t-1} + b_{t-1}) \tag{4}$$

Trend

$$b_t = \beta^* \left(l_t - l_{t-1}\right) + (1-\beta^*) b_{t-1} \tag{5}$$

Seasonal

$$s_t = \frac{\gamma y_t}{\left(l_{t-1} + b_{t-1}\right)} + (1-\gamma) s_{t-m} \tag{6}$$

Forecast

$$y_{t+h|t} = (l_t + b_t h) s_{t-m+h_m^+} \tag{7}$$

where, m: length of time series, h_m^+: (h-1)mod $m + 1$, S_t: seasonal component at time t, γ: a smoothing coefficient (between 0 and 1).

2.1.2 Seasonal ARIMA Model

An ARIMA model contains autoregressive (AR), integrated (I) and moving average (MA) parts which are expressed as ARIMA (p, d, q); where p is autoregressive part, d is integrated part and q is moving average part (Box and Jenkins 1976). AR describes the relationship between present and previous variables in the time series. If $p = 1$, then each variable is a function of only one last variable, i.e.

$$Y_t = c + \varnothing_1 Y_{t-1} + e_t \tag{8}$$

where, Y_t: observed value at time t, Y_{t-1}: previous observed value at time t-1, e_t: random error, c and φ_1: constants. For $p > 1$, other observed values of the series can be included as:.

$$Y_t = c + \varnothing_1 Y_{t-1} + \varnothing_2 Y_{t-2} + \cdots + \varnothing_p Y_{t-p} + e_t \tag{9}$$

The I part of the model denotes the stationarity of the series. For a non-stationary series, differencing has to be done. For linear trend in the time series, first-order differencing is done ($d = 1$), for a quadratic trend, $d = 2$ and so on. The MA part of the model identifies the relationship between the variable and previous q errors. If $q = 1$, each observation is a function of only one previous error i.e.

$$Y_t = c + \theta_1 e_{t-1} + e_t \tag{10}$$

where, c: constant, e_t: random error at time t, e_{t-1}: previous random error at time t–1. For $q > 1$, other errors can be included as:

$$Y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \cdots + \theta_q e_{t-q} + e_t \tag{11}$$

The combined equation for non-seasonal ARIMA model of order (p, d, q) for a standard normal variable (Y_t) is:

$$\varnothing(B)(1 - B)^d Y_t = \theta(B) e_t \tag{12}$$

where, B: backshift operator. To account for seasonality, the ARIMA model is represented by ARIMA $(p,d,q) \times (P,D,Q)_s$ with P, D and Q denoting the seasonal autoregression, integration (differencing), and moving average, respectively.

$$\varnothing_p(B)\Phi_P(B^s)\nabla^d \nabla_s^D Y_t = \theta_q(B)\Theta_Q(B^s) e_t \tag{13}$$

where, Y_t: original time series, e_t: normal independently distributed white noise residual series with mean zero and variance σ^2, Φ and Θ: ARIMA structures between the seasonal observations, φ and θ: non-seasonal ARIMA structure, $\phi_p(B)$ and $\Phi_P(B^s)$: non-seasonal and seasonal autoregressive operators of order p and P, respectively, $\theta_q(B)$ and $\Theta_Q(B^s)$: non-seasonal and seasonal moving average operators of order q and Q, respectively, ∇^d

and ∇_s^D: non-seasonal and seasonal differencing operators of orders d and D. The Seasonal ARIMA procedure involves three major steps –

(a) Model identification: the time series is first analysed for stationarity and normality and accordingly differenced and/or log-transformed. Autocorrelation (ACF) and partial autocorrelation (PACF) functions of the original and differenced series are examined. The best model is identified based on least Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). The AIC or BIC is written in the form [-2logL + kp], where L: likelihood function, p: number of parameters in the model, and k: 2 for AIC and log(n) for BIC (Akaike 1974; Schwarz 2007).
(b) Parameter estimation and diagnostic checking: autoregressive and moving average parameters of the identified model are estimated. Diagnostic checking is carried out from the residuals and accuracy measures to examine the suitability and assumptions of the model.
(c) Forecasting: the selected model is used to forecast the variables for future time periods.

2.1.3 Multi-layer Perceptron

Several AI-based deep learning frameworks are available that can simulate time series data. Of these, MLP is a simple feed-forward layered technique that works on the principle of the backpropagation algorithm (Hecht-Nielsen 1989; Aslam et al. 2020). The data are processed in the forward direction by feeding into input nodes. These are then multiplied by a given weight and passed onto one or more hidden layer nodes. The hidden layer nodes add the weighted inputs, calculate loss or bias and then pass it on through a transfer function to give the result. The output nodes perform the same operations. During the training phase, the weights and biases are adjusted to minimise the errors via the backpropagation algorithm and a comparison is made between target outputs at each output node and output network (Lallahem et al. 2005). The backpropagation algorithm keeps repeating this iteration till the maximum improvement is achieved. Fig. S1 demonstrates a three-layer feed-forward MLP structure. The equation for MLP is given by:

$$y_k = S_1\left(\sum_{j=1}^{J} w_j S_2\left(\sum_{i=1}^{I} w_i x_i + W_j\right) + W_k\right) \tag{14}$$

where, y_k are outputs from the network, x_i are inputs, w_i are weights connecting input and hidden layer nodes, w_j are weights connecting hidden and output layer nodes, I are input nodes, J are hidden nodes, K are output nodes, W_j is bias for jth hidden neuron and W_k is bias for kth output neuron. S_1 and S_2 are activation functions. Commonly, the logistic sigmoid function is used for activation:

$$S(x) = \frac{1}{1 + e^{-x}} S : R \rightarrow]-1, 1[\tag{15}$$

2.1.4 Extreme Learning Machine

ELM was developed to overcome traditional neural networks' high computational cost and time (Huang et al. 2006). In the three-layered structure (single hidden layer feed-forward network) of ELM, weights between inputs and hidden nodes and bias values in the hidden layer are generated at random which are frozen during model training. The weights

between hidden nodes and outputs are calculated using Moore–Penrose generalized inverse of the hidden-output matrix. This process makes ELM much faster than traditional ANNs because there is no iteration in learning. For a training set, ELM is expressed as:

$$\sum_{i=1}^{\hat{N}} \beta_i g\left(w_i x_j + b_i\right) = o_j, j = 1, \ldots, N \tag{16}$$

where, x_i: input, $\hat{N}$: hidden nodes, $w_{i:}$ weight vector from input to ith hidden node, b_i: bias of ith hidden node, β_i: weight vector from ith hidden node to output, o_j: output and g(x): non-linear activation function in hidden layer (Parisouj et al. 2020).

2.1.5 Neural Network Auto-regression

In an NNAR model, lagged values are used as inputs to a neural network. NNAR(p,k) indicates that there are p lagged inputs and k nodes in hidden layer. For seasonal data, last observed values from same season are used as inputs. It is denoted by NNAR(p, P, k)$_{\mathrm{m}}$ which has ($y_{t\text{-}1}$, $y_{t\text{-}2}$, …, $y_{t\text{-}p}$, $y_{t\text{-}m}$, $y_{t\text{-}2\ m}$, …, $y_{t\text{-}Pm}$) as inputs and k neurons in hidden layer. Unlike SARIMA, NNAR does not require stationarity of time series. Mathematically NNAR is:

$$y_t = f\left(y_{t-1}\right) + \varepsilon_t \tag{17}$$

where, f: neural network, y_{t-1}: vector containing lagged values of series and ε_t: error series (assumed to be homoscedastic) (Hyndman and Athanasopoulos 2018).

2.2 Study Area and Data Used

The five forecasting models were applied on groundwater records of monitoring wells in the National Capital Territory (NCT) of Delhi. The NCT of Delhi in North India lies between 28° 24′ 15″ N to 28° 53′ 00″ N and 76° 50′24″ E to 77° 20′ 30″ E. It covers an area of 1483 km^2, 75% of which is urbanized (Central Ground Water Board 2016). It is divided into 11 districts. Much of Delhi's rain occurs during the monsoon months – July to September, bringing high humidity levels. Delhi's geological formations vary from Quartzite to Older and Younger Alluvium making the aquifer geology complex (CGWB 2021a).

The Central Ground Water Board (CGWB) has over a 100 monitoring stations spread over Delhi's alluvial and quartzitic area. Historical records of groundwater levels for monitoring stations in NCT of Delhi were obtained from CGWB for winter (January), pre-monsoon (May), monsoon (August) and post-monsoon (November) seasons. Mann–Kendall test on pre-monsoon and post-monsoon groundwater levels (in mbgl) for Delhi at the district level showed an increasing trend indicating that there has been a decline in the groundwater depth in the last two decades (Sarma and Singh 2021b).

2.3 Data Pre-processing and Methodology of Study

To demonstrate the comparison between the time series models, data from three wells were chosen because of their declining groundwater level – Haiderpur (GW1), Bhatti (GW2) and Kitchner Road (GW3). Locations of these wells are presented in Fig. 1. The time period for study was selected depending on data availability – GW1 (1999–2019), GW2 (1996–2019) and GW3 (1983–2017). The datasets were analysed for completeness

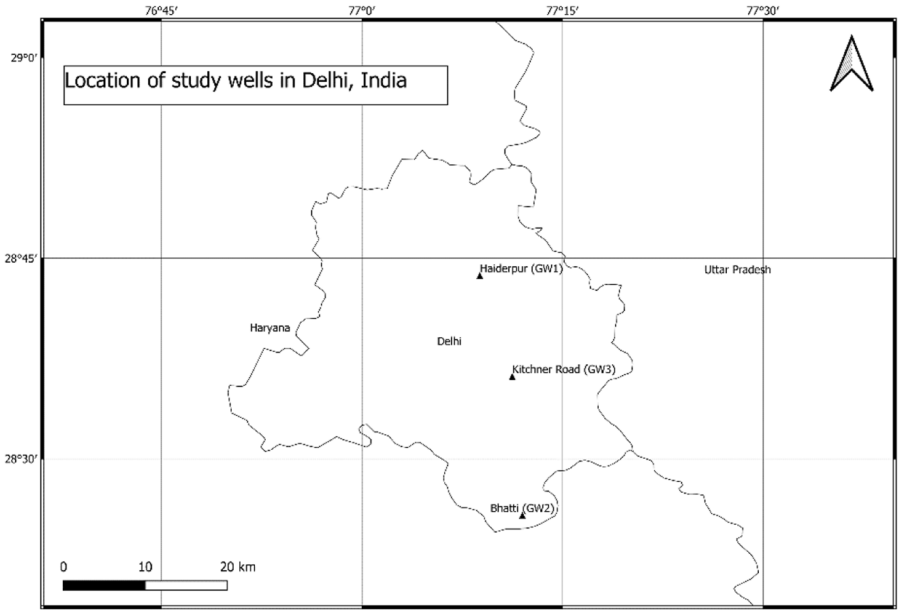


Fig. 1 Locations of selected monitoring wells for the study – Haiderpur (GW1), Bhatti (GW2) and Kitchner Road (GW3)

and a small number of missing values were imputed using the Multivariate Imputation by Chained Equations (MICE) package in R software (van Buuren and Groothuis-Oudshoorn 2011). Multiple imputation has an advantage over single imputation methods in that correlations error obtained from the imputations are not overestimated due to the inclusion of uncertainty because of missing data (Lee and Carlin 2010; Gao et al. 2018). The dataset was divided into two subsets for applying the time series models–80% for training and 20% for testing. To avoid larger values from overriding smaller ones and prevent saturation of hidden nodes, the dataset was normalized between 0 and 1 for MLP and ELM (Eq. (18)).

$$x_n = \frac{(x_i - x_{min})}{(x_{max} - x_{min})} \tag{18}$$

where, x_n: normalized data, x_i: actual value, x_{min}: minimum value and x_{max}: maximum value in each dataset (Shirmohammadi et al. 2013). The HWES, NNAR, SARIMA, MLP and ELM algorithms were run on the training data using their respective packages in R software – *forecast* (Hyndman et al. 2008, 2019; Hyndman and Athanasopoulos 2018), *astsa* (Shumway and Stoffer 2016) and *nnfor* (Crone and Kourentzes 2010; Kourentzes et al. 2014). The complete methodology of the study is depicted in Fig. 2.

2.4 Evaluation of Model Performance

After fitting the training data, each model was used to forecast the groundwater level for the testing period. To assess model performance, accuracy measures–root mean square

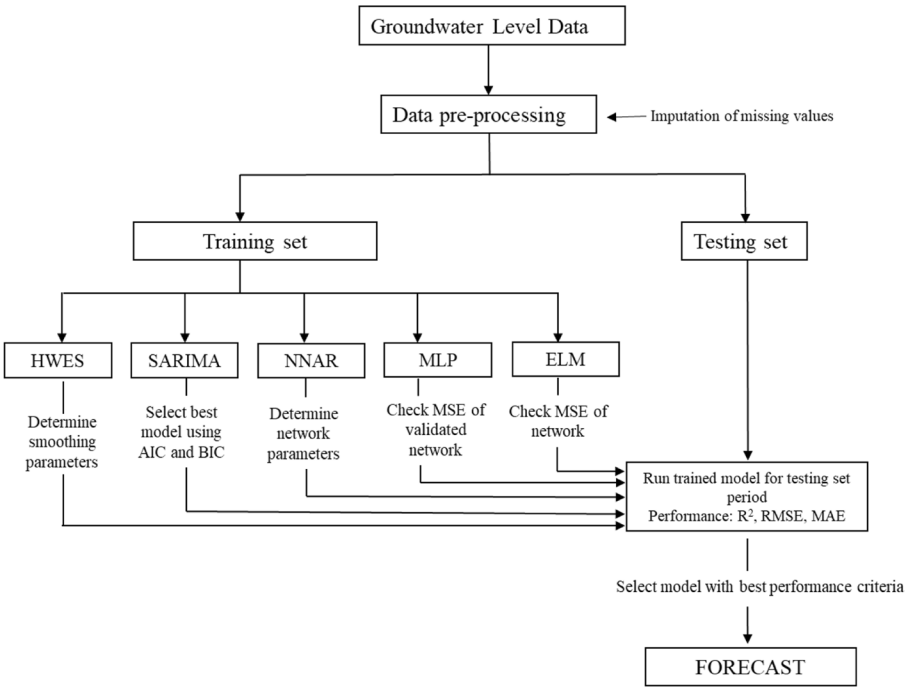


Fig. 2 Methodology of study

error (RMSE), mean absolute error (MAE) and coefficient of determination (R^2) were compared. The model with the least RMSE and MAE and maximum R^2 during the training and testing phases was determined as the best model for forecasting groundwater levels in the study area.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(O_i - P_i)^2} \tag{19}$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|(O_i - P_i)| \tag{20}$$

$$R^2 = \left[\frac{\sum_{i=1}^{n}(O_i - \underline{O})(P_i - \underline{P})}{\sqrt{\sum_{i=1}^{n}(O_i - \underline{O})^2 \sum_{i=1}^{n}(P_i - \underline{P})^2}}\right]^2 \tag{21}$$

where, n: number of data points, O_i: observed variables with mean $\underline{O}$ and P_i: predicted variables with mean $\underline{P}$ (Choubin and Malekian 2017; Yan and Ma 2016).

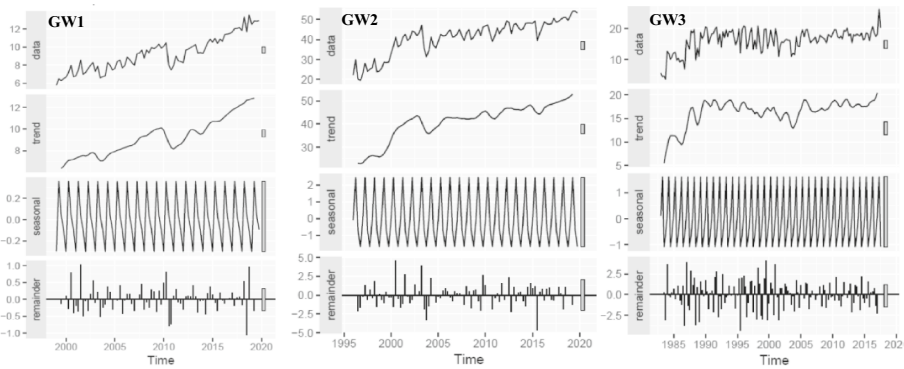


Fig. 3 Seasonal decomposition of groundwater level time series of GW1, GW2 and GW3 into trend, seasonal and random components

3 Results and Discussion

3.1 Time Series Decomposition

The time series of the three wells were first decomposed into its trend, seasonal and residual components in R software (Fig. 3). Since the data are in metres below ground level (mbgl), all 3 wells showed increasing trend, implying continuous groundwater depletion during the time period of study. The trend ranged over 7 mbgl for GW1, 30 mbgl for GW2 and 15 mbgl for GW3. The seasonal component has an oscillatory amplitude of 0.3 mbgl for GW1, 3 mbgl for GW2 and 2 mbgl for GW3. The seasonal component has a peak in November and trough in May, according to the rainfall patterns in Delhi. Much of the rainfall is received during the monsoon season (July – September) which recharges the groundwater during the post-monsoon months (October – November). During January to June, scanty rainfall is unable to compensate for the groundwater abstraction, thus leading to a lower recharge in May. The residual component fluctuates between ± 1 mbgl for GW1, ± 5 mbgl for GW2 and ± 3 mbgl for GW3. This decomposition showed that for all 3 study wells, the trend has the largest component, followed by the residual and seasonal parts.

3.2 Holt-winters' Exponential Smoothing

The Holt-Winters' model predicts a future value by combining the influences from the level, trend and seasonality. Each component has an associated smoothing parameter that provides information about its influence on the model's predictions (Table 1). The value

Table 1 Smoothing parameters from HWES with trend and additive seasonal component

	Smoothing parameters		
	α	β*	γ
GW1	0.8191	0.0321	0.2680
GW2	0.7383	0.0000	0.6628
GW3	0.4267	0.0897	0.2982

of α is close to 1 for GW1 and GW2, indicating that prediction of a current observation is mostly based on the immediate past observations. For GW3, lower value of α implies that older observations are weighted more than the most recent ones. All three study wells had β* close to or equal to 0 indicating that the trend does not change much over time and remains fairly constant during each prediction. GW1 and GW3 have γ values close to 0 indicating that like trend, seasonality also does not change. However, for GW2, a larger γ value implies a strong seasonal component. Thus, for GW2, seasonal component of the current observation is based on the seasonality of the most recent observations.

3.3 Seasonal ARIMA

The autocorrelation (ACF) and partial autocorrelation (PACF) plots for the time series of GW1, GW2 and GW3 were inspected to check the stationarity. The plots did not tail off to zero indicating that the series were not stationary and required appropriate differencing. First order non-seasonal differencing was applied on GW1 and GW3 while GW2 required first order seasonal. This corroborated the results from HWES as GW2 showed a strong seasonal component while GW1 and GW3 did not have much influence from seasonality. The Augmented Dickey-Fuller (ADF) test was applied before and after differencing and the p-value was compared. The p-value before differencing was computed as 0.4451, 0.1118 and 0.0419 for GW1, GW2 and GW3 respectively. After the appropriate differencing, the p-value as determined from the ADF test was 0.01, rejecting the null hypothesis that the series has a unit root and is thus stationary.

With the appropriate differencing, the value of d or D of the SARIMA model was determined. Different combinations of p, P, q and Q were tested on trial-and-error method. The ACF and PACF plots of each of these models were inspected and model selection criteria AIC and BIC were compared (Table S1). The best model was selected based on the minimum AIC and BIC values. The final models selected were $(1,1,1)(1,0,0)_4$, $(1,0,0)(0,1,1)_4$ and $(1,1,1)(1,0,1)_4$ for GW1, GW2 and GW3 respectively.

3.4 MLP, ELM and NNAR

The MLP networks were trained in R using a validation argument and specified number of lags to prevent overfitting the training data (Fig. 4). The grey inputs represent

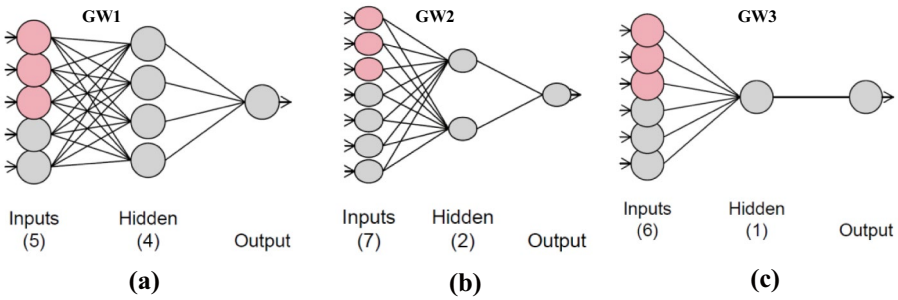


Fig. 4 MLP network for **a** GW1 (5 inputs, 1 hidden layer, 4 nodes) **b** GW2 (7 inputs, 1 hidden layer, 2 nodes) and **c** GW3 (6 inputs, 1 hidden layer, 1 node)

Table 2 ELM and NNAR model descriptions for the study wells

	ELM network	**Univariate lags**
GW1	Fit with 62 hidden nodes	1,3
GW2	Fit with 69 hidden nodes	1,2,3,4
GW3	Fit with 100 nodes	1,2,3
	NNAR model	**Description**
GW1	$NNAR(1,1,2)_4$	2–2-1 network with 9 weights
GW2	$NNAR(1,1,2)_4$	2–2-1 network with 9 weights
GW3	$NNAR(9,1,5)_4$	9–5-1 network with 56 weights

autoregressive lags and pink inputs represent seasonality. The mean square error (MSE) for the trained MLP networks were 0.0022, 0.0036 and 0.0127 for GW1, GW2 and GW3 respectively. The ELM networks were trained in R using the lasso regression method. The time taken for training was significantly lesser than the MLP networks. The mean square error (MSE) for the trained ELM networks were 0.0072, 0.0062 and 0.0156 for GW1, GW2 and GW3 respectively. The NNAR models were applied on the training data with Box-Cox transformation (lambda = 0). This transformation ensured that the residuals were homoscedastic. The descriptions of the resultant ELM and NNAR networks are presented in Table 2.

3.5 Comparison of Model Performance

The accuracy measures were compared for the models for each study well (Table 3). For GW1 and GW2, MLP performed the best for both the training and testing phases with highest R^2 and lowest RMSE and MAE. MLP best fitted the observed values with the predicted values during training and testing (Fig. 5). For GW3, HWES had the highest RMSE and MAE among all models. NNAR performed very well during training but gave unrealistic forecasts during testing. There may have been overfitting of data, indicated by the large parameters in the network, i.e. 9–5-1 network with 56 weights. The MLP network yielded low R^2 values for both training and testing. MLP network for GW3 had only 1 hidden node (Fig. 4) and compared to GW1 and GW2, MSE for GW3 was higher. ELM presented very high RMSE and MAE for both training and testing. SARIMA had slightly better R^2 but MLP had the lowest errors. Thus, MLP was considered the best performing model for GW3 as well.

The best model MLP was used to forecast the groundwater level for the year 2025. MLP forecasts for May 2025 showed that groundwater level will fall by 2 mbgl and 21 mbgl below the 2019 level for GW1 and GW2 respectively and 3 mbgl below the 2017 level for GW3.

3.6 Discussion

The results from applying the five models suggest that with validation in the training set, MLP can train the model with R^2 as high as 0.914. For GW1 and GW2, SARIMA and

Table 3 Comparison of accuracy measures for selected models

	Training			Testing		
	R^2	RMSE	MAE	R^2	RMSE	MAE
GW1						
HWES	0.746	0.638	0.455	0.729	0.597	0.469
SARIMA	0.800	0.617	0.479	0.379	1.813	1.616
MLP	0.914	0.047	0.035	0.733	0.069	0.052
ELM	0.731	0.085	0.062	0.721	0.112	0.093
NNAR	0.793	0.561	0.431	0.516	2.266	2.111
GW2						
HWES	0.882	2.557	1.822	0.708	5.377	5.041
SARIMA	0.912	2.448	1.676	0.671	5.033	4.612
MLP	0.910	0.060	0.041	0.710	0.139	0.128
ELM	0.852	0.079	0.058	0.705	0.112	0.095
NNAR	0.873	2.541	1.924	0.633	5.166	4.595
GW3						
HWES	0.271	3.060	2.323	0.284	3.071	2.803
SARIMA	0.558	2.645	1.989	0.246	1.839	1.351
MLP	0.436	0.113	0.088	0.195	0.104	0.086
ELM	0.354	2.793	2.021	0.266	2.282	1.902
NNAR	0.931	0.904	0.659	0.017	2.263	1.475

HWES also gave high R^2 values during training. For the testing period, the R^2 tends to decrease slightly for MLP and HWES and dramatically for SARIMA. Thus, fitting precision of a model may not necessarily imply accurate forecasting which makes evaluating the model forecasts against a testing subset a crucial step in the modeling process. ELM gave the lowest R^2 for GW1 and GW2 during training. NNAR had overall low efficiency which was also observed in the study by Aguilera et al. (2019). These models may perform better by selecting the appropriate activation functions, lagged variables as inputs or numerical procedure (Faraway and Chatfield 1998). MLP, ELM, HWES and SARIMA did not perform well on the GW3 dataset. Even though the dataset for GW3 was large with 139 observations over 1983–2017, there was overall low performance of the models. Groundwater levels are influenced by factors like rainfall, soil properties, surface water and abstraction (Lee et al. 2019) and during interpretation of model results, it may be pertinent to analyse these influencing factors as well.

MLP forecast for 2025 presents an alarming result for GW2, a monitoring well located in the southern part of the study area (Fig. 1). Declining trends in the South district of Delhi have been reported by the CGWB, particularly due to over-exploitation (CGWB 2016, 2021b). Results of aquifer response modeling for Delhi using MODFLOW also showed a similar decline for GW2 (Bhatti) (CGWB 2016), necessitating an urgent and effective groundwater management plan in that region.

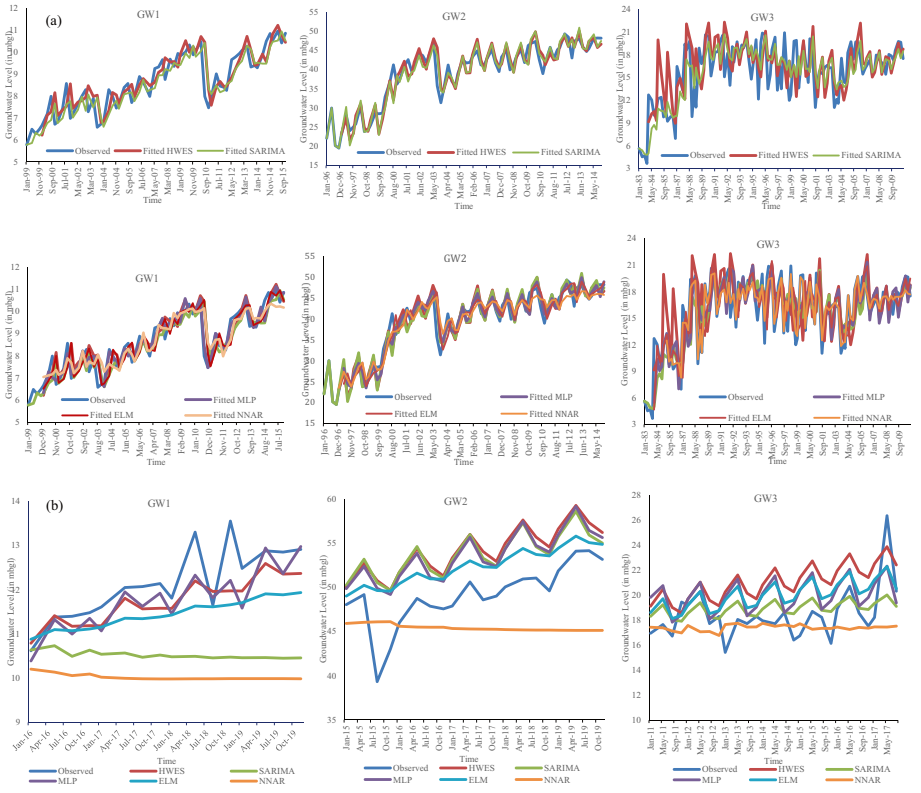


Fig. 5 Fitting the observed and simulated values for **a** training and **b** testing

4 Conclusion

This study used univariate time series forecasting methods HWES, Seasonal ARIMA, MLP, ELM and NNAR for prediction of groundwater levels in Delhi, India. Decomposition of the time series of three monitoring wells showed increasing trend, indicating that the groundwater level has continuously declined over the study period. The parameters of HWES and SARIMA were calculated and MLP, ELM and NNAR models were trained and validated. The accuracy of the models was assessed based on the R^2, RMSE and MAE. MLP had the least values of RMSE and MAE and highest R^2 for two wells during training and testing, indicating that the MLP approach was most accurate in forecasting the groundwater levels. For the third well, MLP had a slightly lower R^2 but least RMSE and MAE and was also concluded to be better than HWES, SARIMA, ELM and NNAR for making predictions.

Such univariate forecasting studies are particularly suitable for regions where large hydro-climatological data are unavailable. MLP was used to make further predictions for May 2025 for all three wells. Results indicate that groundwater level in the three wells will decline by 2–21 mbgl. This study has helped identify the areas that require urgent groundwater management decisions. Further research may be done on adding more arguments in the model algorithm to give the lowest errors. It is recommended that more comparison

studies should be made for other regions in India so that the most accurate models help in identifying the areas where groundwater is declining rapidly.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11269-022-03173-6.

Acknowledgements The authors acknowledge the support of the Central Ground Water Board, India and thank the CGWB and India-WRIS for providing the data for this study. The authors also thank the Delhi Technological University, New Delhi for providing the facilities for this study.

Author Contributions Conceptualization and methodology: RS, SKS; Analysis and investigation: RS; Supervision: SKS; Writing – original draft preparation: RS; Writing – review and editing: RS, SKS.

Availability of Data and Material The data used in this study have been obtained from the Central Ground Water Board, India and are available in the manuscript.

Declarations

Ethical approval Not applicable.

Consent to Participate Not applicable.

Consent to Publish The authors hereby grant all rights of publication of the manuscript to the publisher.

Competing Interests The authors have no relevant financial or non-financial interests to disclose.

References

Aguilera H, Guardiola-Albert C, Naranjo-Fernández N, Kohfahl C (2019) Towards flexible groundwater-level prediction for adaptive water management: using Facebook's Prophet forecasting approach. Hydrol Sci J 64(12):1504–1518. https://doi.org/10.1080/02626667.2019.1651933

Akaike H (1974) A New Look at the Statistical Model Identification. IEEE Trans Autom Control 19(6):716–723. https://doi.org/10.1109/TAC.1974.1100705

Aslam M, Lee JM, Hong S (2020) A multi-layer perceptron based deep learning model to quantify the energy potentials of a thin film a-Si PV system. Energy Rep 6:1331–1336. https://doi.org/10.1016/j.egyr.2020.11.025

Barzegar R, Asghari Moghaddam A, Adamowski J, Fijani E (2017) Comparison of machine learning models for predicting fluoride contamination in groundwater. Stochastic Environ Res Risk Assess 31(10):2705–2718. https://doi.org/10.1007/s00477-016-1338-z

Box GEP, Jenkins GM (1976) Series analysis forecasting and control. Prentice-Hall Inc., London

Brown RG (1959) Statistical forecasting for inventory control. McGraw-Hill, New York

Central Ground Water Board (2016) Aquifer mapping and ground water management plan of NCT Delhi. New Delhi. http://cgwb.gov.in/AQM/AQM-Reports.html. Accessed 20 Oct 2021

Central Ground Water Board (2021) Groundwater Yearbook 2019-20. New Delhi. http://cgwb.gov.in/GW-Year-Book-State.html. Accessed 15 Nov 2021

Central Ground Water Board (2021b) Report: Dynamic ground water resources of NCT Delhi as on March 2020. New Delhi. http://cgwb.gov.in/Dynamic-GWResources.html. Accessed 20 Nov 2021

Chen C, He W, Zhou H, Xue Y, Zhu M (2020) A comparative study among machine learning and numerical models for simulating groundwater dynamics in the Heihe River Basin. Northwestern China Sci Rep 10:3904. https://doi.org/10.1038/s41598-020-60698-9

Choubin B, Malekian A (2017) Combined gamma and M-test-based ANN and ARIMA models for groundwater fluctuation forecasting in semiarid regions. Environ Earth Sci 76(15). https://doi.org/10.1007/s12665-017-6870-8

Crone SF, Kourentzes N (2010) Feature selection for time series prediction - A combined filter and wrapper approach for neural networks. Neurocomputing 73(10–12):1923–1936. https://doi.org/10.1016/j.neucom.2010.01.017

Faraway J, Chatfield C (1998) Time series forecasting with neural networks: A comparative study using the airline data. J R Stat Soc, C: Appl Stat 47(2):231–250. https://doi.org/10.1111/1467-9876.00109

Gao Y, Merz C, Lischeid G, Schneider M (2018) A review on missing hydrological data processing. Environ Earth Sci 77:47. https://doi.org/10.1007/s12665-018-7228-6
Hecht-Nielsen R (1989) Theory of the backpropagation neural network. In International 1989 Joint Conference on Neural Networks. Washington, DC, USA
Holt CC (1957) Forecasting trends and seasonals by exponentially weighted averages. In Carnegie Institute of Technology, Pittsburgh ONR memorandum no. 52
Huang GB, Zhu QY, Siew CK (2006) Extreme learning machine: theory and applications. Neurocomputing 70(1):489–501. https://doi.org/10.1016/j.neucom.2005.12.126
Hyndman RJ, Koehler AB, Ord JK, Snyder RD (2008) Forecasting with Exponential Smoothing: The State Space Approach. Springer
Hyndman R et al (2019) Forecast: forecasting functions for time series and linear models. R package version 8.7. http://pkg.robjhyndman.com/forecast. Accessed 15 Oct 2021
Hyndman RJ, Athanasopoulos G (2018) Forecasting: principles and practice. Melbourne, Australia: OTexts. https://otexts.org/fpp2/. Accessed 20 Oct 2021
Kalteh AM (2019) Modular wavelet-extreme learning machine: A new approach for forecasting daily rainfall. Water Resour Manag 33:3831–3849. https://doi.org/10.1007/s11269-019-02333-5
Kourentzes N, Barrow DK, Crone SF (2014) Neural network ensemble operators for time series forecasting. Expert Syst Appl 41(9):4235–4244. https://doi.org/10.1016/j.eswa.2013.12.011
Lallahem S, Mania J, Hani A, Najjar Y (2005) On the use of neural networks to evaluate groundwater levels in fractured media. J Hydrol 307(1–4):92–111. https://doi.org/10.1016/j.jhydrol.2004.10.005
Lee KJ, Carlin JB (2010) Multiple imputation for missing data: fully conditional specification versus multivariate normal imputation. Amer J Epidemiol 171:624–632
Lee S, Lee KK, Yoon H (2019) Using artificial neural network models for groundwater level forecasting and assessment of the relative impacts of influencing factors. Hydrogeol J 27(2):567–579. https://doi.org/10.1007/s10040-018-1866-3
Mirzavand M, Ghazavi R (2015) A stochastic modelling technique for groundwater level forecasting in an arid environment using time series methods. Water Resour Manag 29(4):1315–1328. https://doi.org/10.1007/s11269-014-0875-9
Mozaffari S, Javadi S, Moghaddam HK, Randhir TO (2022) Forecasting groundwater levels using a hybrid of support vector regression and particle swarm optimization. Water Resour Manag. https://doi.org/10.1007/s11269-022-03118-z
Narayanan P, Basistha A, Sarkar S, Kamna S (2013) Trend analysis and ARIMA modelling of pre-monsoon rainfall data for western India. C R Geosci 345(1):22–27. https://doi.org/10.1016/j.crte.2012.12.001
Natarajan N, Sudheer C (2020) Groundwater level forecasting using soft computing techniques. Neural Comput Appl 32:7691–7708. https://doi.org/10.1007/s00521-019-04234-5
Noori AR, Singh SK (2021) Spatial and temporal trend analysis of groundwater levels and regional groundwater drought assessment of Kabul. Afghanistan Environ Earth Sci 80:698. https://doi.org/10.1007/s12665-021-10005-0
Parisouj P, Mohebzadeh H, Lee T (2020) Employing machine learning algorithms for streamflow prediction: A case study of four river basins with different climatic zones in the United States. Water Resour Manag 34:4113–4131. https://doi.org/10.1007/s11269-020-02659-5
Poursaeid M, Poursaeid AH, Shabanlou SA (2022) A comparative study of artificial intelligence models and a statistical method for groundwater level prediction. Water Resour Manag 36:1499–1519. https://doi.org/10.1007/s11269-022-03070-y
Rahaman M, Thakur B, Kalra A, Ahmad S (2019) Modeling of GRACE-derived groundwater information in the colorado river basin. Hydrol 6(1):19. https://doi.org/10.3390/hydrology6010019
Sakizadeh M, Mohamed MMA, Klammler H (2019) Trend analysis and spatial prediction of groundwater levels using time series forecasting and a novel spatio-temporal method. Water Resour Manag 33(4):1425–1437. https://doi.org/10.1007/s11269-019-02208-9
Salvadore E, Bronders J, Batelaan O (2015) Hydrological modelling of urbanized catchments: A review and future directions. J Hydrol 529(1):62–81. https://doi.org/10.1016/j.jhydrol.2015.06.028
Sarma R, Singh SK (2021a) Temporal variation of groundwater levels by time series analysis for NCT of Delhi, India. In: Mehta YA, Carnacina I, Kumar DN, Rao KR, Kumari M (eds) Advances in Water Resources and Transportation Engineering. Lecture Notes in Civil Engineering (149). Springer, Singapore. https://doi.org/10.1007/978-981-16-1303-6_15
Sarma R, Singh SK (2021b) Simulating contaminant transport in unsaturated and saturated groundwater zones. Water Environ Res 93(9):1496–1509. https://doi.org/10.1002/wer.1555
Schwarz G (2007) Estimating the dimension of a model. Ann Stat 6(2):461–464. https://doi.org/10.1214/aos/1176344136

Shirmohammadi B, Vafakhah M, Moosavi V, Moghaddamnia A (2013) Application of several data-driven techniques for predicting groundwater level. Water Resour Manag 27(2):419–432. https://doi.org/10.1007/s11269-012-0194-y

Shumway RH, Stoffer DS (2016) Time series analysis and its applications With R Examples EZ Edition. Springer Texts in Statistics. Springer Series in Statistics, EZ Edition, New York

Singh A (2014) Groundwater resources management through the applications of simulation modeling: A review. Sci Total Environ 499:414–423. https://doi.org/10.1016/j.scitotenv.2014.05.048

Takafuji EHDM, Rocha MMD, Manzione RL (2019) Groundwater level prediction/forecasting and assessment of uncertainty using SGS and ARIMA models: A case study in the Bauru Aquifer System (Brazil). Nat Resour Res 28(2):487–503. https://doi.org/10.1007/s11053-018-9403-6

van Buuren S, Groothuis-Oudshoorn K (2011) MICE: Multivariate imputation by chained equations in R. J Stat Softw 45(3):1–67. https://www.jstatsoft.org/v45/i03/

Winters PR (1960) Forecasting sales by exponentially weighted moving averages. Manage Sci 6(3):324–342. https://doi.org/10.1287/mnsc.6.3.324

Yan Q, Ma C (2016) Application of integrated ARIMA and RBF network for groundwater level forecasting. Environ Earth Sci 75(5):1–13. https://doi.org/10.1007/s12665-015-5198-5

Yang Q, Wang Y, Zhang J, Delgado J (2017) A comparative study of shallow groundwater level simulation with three time series models in a coastal aquifer of South China. Appl Water Sci 7(2):689–698. https://doi.org/10.1007/s13201-015-0282-2

Yang ZP, Lu WX, Long YQ, Li P (2009) Application and comparison of two prediction models for groundwater levels: A case study in Western Jilin Province. China J Arid Environ 73(4–5):487–492. https://doi.org/10.1016/j.jaridenv.2008.11.008

Zhang Z, Zhang Q, Singh VP (2018) Univariate streamflow forecasting using commonly used data-driven models: literature review and case study. Hydrol Sci J 63(7):1091–1111. https://doi.org/10.1080/02626667.2018.1469756

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

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760762

A Comprehensive Study on the Role of Machine Learning in Hybrid Biometric Recognition

Shipra Shand
Department Of Software Engineering
Delhi Technological University
New Delhi, India
meetshipra43@gmail.com

Rahul
Department Of Software Engineering
Delhi Technological University
New Delhi, India
rahul@dtu.ac.in

***Abstract*–There is a potential issue with unimodal biometric recognition systems. If the system is only able to use a single type of biometric, such as fingerprints, there is a risk that the system could be fooled by a fake fingerprint. As a result, this research study investigates how a hybrid biometric recognition system (HBRS) can overcome the problem with biometrics that rely on a single attribute for recognition. A HBRS is a system that uses more than one type of biometric recognition technology. Machine learning (ML) is used in biometric recognition to improve the accuracy of the recognition system. Therefore, this research study explore the role of machine learning in this identification process, as well as several machine learning algorithms that aid in the construction of HBRS systems, to see which methods provide the best accuracy.**

Keywords– Machine Learning(ML), Hybrid Biometric Recognition System(HBRS), Supervised Learning(SL), Unsupervised Learning(USL).

I. INTRODUCTION

Machine learning can be considered as a method or other process of data analysis that generally automates in creating analytical models[1]. It is mainly a branch about artificial intelligence depending upon the thought that systems might learn and gain knowledge from data, identification of patterns, including that making decisions with minimal human intervention. ML procedures support manufacturers in reducing procedure driven losses, enhancing capacity through optimizing the production-oriented procedure including that it also minimizes prices through conducting predictive maintenance. Critical roles including the importance of this ML in ***HBRS*** will be briefly discussed in this study. The Hybrid biometric recognition related approach is mostly necessary for identification including that verification in single or too many systems. This approach of hybrid biometric is also supported in such circumstances where a particular biometric feature is just not optimal regarding particular user's groups. In modern times this biometric recognition is a popular pattern for not just identification purposes it is also utilized in several organizations, government related works including that it is also vastly utilized in crime prevention purposes [27]. With the support of hybrid biometric recognition basically as facial features including fingerprint recognition, it would facilitate the process regarding law enforcement officials in terms of spotting and tracking dangerous criminals. Biometric recognition systems vary based upon the human's own personal traits including their characteristics. Machine learning can also help in studying biometric topographies in order to stimulate a separate identification learning and gaining knowledge-oriented activities. Machine learning uses a vast estimation of structured data in case that it might be able to produce outcomes or else provide several predictions depending on those provided data. In the present scenario, this machine learning has been utilized in identification fingerprint recognition, facial recognition, finger vein procedures. Mainly there are three approaches to ML and these approaches are SL, RL, and USL.

II. RELATED WORK

In their study, Khan, Amreen, and Abhishek Bhatt[4] constructed an HBRS that utilized fingerprint and finger vein as biometric traits. Using RF and SDAE classifiers, they were able to attain a 97 percent accuracy rate.

In their study, Hossain, M.A., and Al Hasan[5] improved cloud security by combining fingerprint biometrics with encryption methods.

In their study, Jacob and I. Jeena[27] constructed a BRS that uses the retina as a biometric characteristic. They used a new algorithm called capsule network and the datasets CASIA-Iris and face-95 to obtain a 99 percent accuracy.

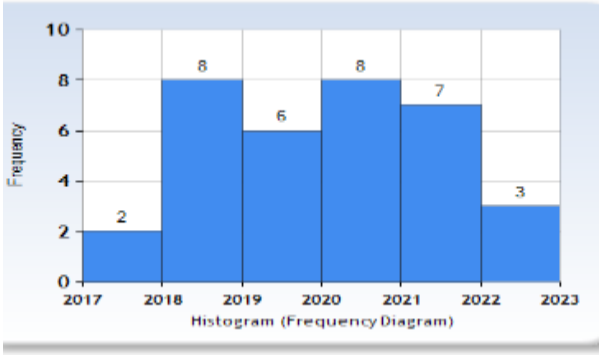


Fig.1. Depicts the histogram of the number of papers studied.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

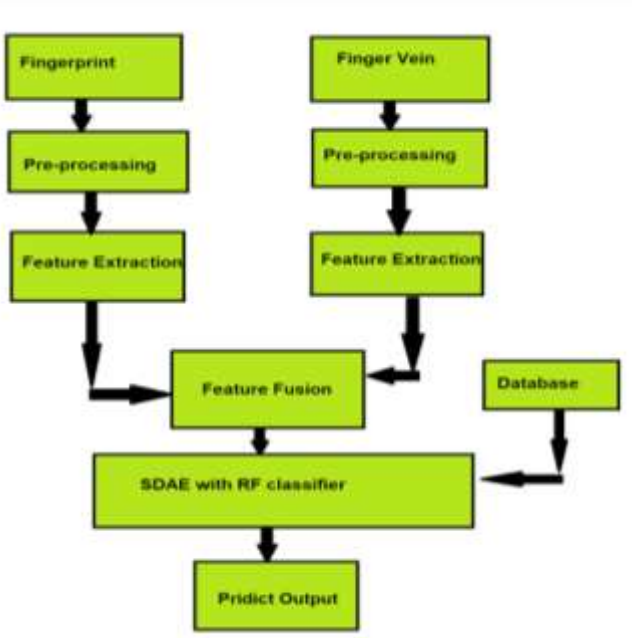


Fig.2. This diagram depicts the methods used by Khan, Amreen, and Abhishek Bhatt[4].

Mady, Huda, and Shadi MS Hilles[31] developed a system that recognises faces based on many aspects such as expressions, body position, and so on. Random forest (RF) was the best strategy they used, with a score of 97.6 percent accuracy. RF is made up of a number of Decision-Trees(DT) [32].

Ponnusamy and Vasaki [33] suggested an HBAS in which they discovered that MNB (Multinomial-Naive-Bayes) performed better than other classifiers. This MNB is based on Bayes' theory [34], which has been around for a long time.

TABLE I. THE OVERVIEW OF PAPERS EXAMINED.

Author	Year of Publication	Algorithm Used	Dataset Used	Remarks
Patro, Kiran Kumar, [22]	2017	SVM, KNN, ANN	MIT-BIH, ECG-ID	SVM gives an accuracy of 93.709%
Ahmadi, N. and Akbarizadeh,G[12]	2017	MLPNN, PSO	UCI, CASIA V3	Performs better as compared to other BR technique
Kumar, R.H., Haripriya, D. and Murikipudi, S.[7]	2018	GSM, Cloud, Android	Attendance of student taken as input	Sends SMS to parents when students attendance is less than 75%.
Bora, A., Vajpai, J. and Gaur, S[9]	2018	LPC; MFCC; MFLPC; PCA	Voice is taken as a dataset.	LPC-with MFCC gives optimum accuracy.
Oloyede, M.O., Hancke, G.P. and Myburgh, H.C[10]	2018	SOA , CNN	Standard dataset of face images.	The proposed algorithm increases the efficiency of recognition.
Prabu, S., Lakshmanan, M. and Mohammed,[3]	2019	HAF, ELBP, SIFT, NN, BN	CASIA Image Dataset	Accuracy of 98.5% was achieved.
Mohsin, A.H, Zaidan.[6]	2019	PSO, AES, RFID,CIA	Finger Vein Images	Accuracy of 100% in FV identification.
Arshad, Habiba, et al.[38]	2020	D-CNN, FEcS	AVAMVG gait, CASIA A, B and C.	99.8, 99.7, 93.3 and 92.2%, were the accuracies of the respective datasets
M. Ilham Rizqyawan, Ulfah Nadiya [14]	2021	DT, SVMRF, Ada-Boost	Facial dataset taken as input	SVM gives the best accuracy of 98%.
Zhang, Liping, et al.[18]	2021	Attention-Network, Features - Network	Facial -Dataset	Proposed algorithm is better than others.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

III. METHODOLOGY

A. Concept of Hybrid Biometric Recognition

"Hybrid Biometric Platform" can be considered as a multimodal biometrics process that generally helps in face, finger vein, and fingerprint identification[10][16]. ***Biometric identification process*** or else system helps in identifying humans through their own personal traits basically as characteristics of their own fingerprints, veins, face, DNA, palm prints or else eyes. The authenticity along with difficulty of this hybrid Biometric recognition system varies based upon the human's own personal traits including their characteristics. Moreover, it also varies based upon measuring the sphere, multi-modal identification or else hybrid biometrics system that helps in measuring and identifying more than single trait that is becoming more favored [3]. In case single measurement losses in providing ample information, the other part serves basically as a backup in order to support appropriate identification or else decision making. This hybrid biometric platform mainly gives peace of mind, gives flexibility along with that insurance. This system naturally enables tailoring the deployment model depending upon the different necessities of the consumers; or else end users and simply switch biometric engines instead of installing any kind of latest software. For identification, the HBRS passes through several phases.

The stages of a generic HBRS, as we know, are [4]:

1. Pictures with various attributes to be input.
2. Pictures have to be pre-processed.
3. Extraction of features from the input data.
4. Fusion of features.
5. Input to a classification algorithm that is appropriate.
6. making a decision.

We'd additionally want to mention the performance parameters that are commonly used to evaluate the proposed system's performance. These are the following[4]:

1. Accuracy=(TP+TN)/(TP+TN+FP+FN).
2. Recall=TP/(TP+FN).
3. Precision=TP/(TP+FP).
4. False Alarm Rate(FAR)=FP/(TN+FP).
5. False Rejection Rate(FRR)=FN/(TP+FN).
6. Equal Error Rate(EER)=(FAR+FRR)/2.

where TP-true positive samples, TN-true negative samples, FP-false positive samples and FN- false negative samples. The parameter accuracy determines how accurate the proposed algorithm is. Recall is the ratio of relevant items that are retrieved to the total number of relevant items. Precision is a measure of how close two or more measurements of the same quantity are to each other. The higher the precision, the closer the measurements are.

In machine learning, the false alarm rate is the rate at which incorrect predictions are made. False rejection rate is the probability that an algorithm will incorrectly reject a truly positive instance. The equal error rate (EER) is the error rate where the false positive rate and the false negative rate are equal. Using the above performance indicators, we can simply determine and prove the suggested method's performance and accuracy.

Utilization of this hybrid biometric has several benefits such as just single integration gives access towards an enterprise organized hybrid biometrics-oriented system. Biometric related identifiers are mainly the measurable characteristics, distinctive that are utilized in labeling including that describing individuals. Biometrics hybrid-oriented system technique that naturally is amalgamation of more than single biometric options in terms of creating verification along with authentication; might be utilized in terms of overcoming several limitations of this mono modal biometrics [4]. This hybrid biometric recognition procedure is now a popular method including that for several security purposes along with applications Iris including Fingerprint recognition method utilization are increasing day by day. Biometrics can be considered as a path of measuring a human's physical characteristics in terms of identifying their own identity [3]. These generally comprise physiological oriented traits basically as fingerprints including eyes or else behavioral characteristics mainly as the different path that someone fulfills a security authentication-oriented puzzle.

As biometrics are tough to prove it is fake or else stealing, therefore, for high security purposes and assurance this system is basically utilized. The enhancement in security level because of the utilization of these recognition methods in comparison to traditional identifying methods is supporting the quick advancement of this system. The utilization of Iris or else Fingerprint recognition process can be considered as more reliable including that authentic method[4].

B. Role of Machine Learning in Hybrid Biometric Recognition

Machine learning can also help in studying biometric topographies in order to stimulate a separate identification learning and gaining knowledge-oriented activities. In addition, these approaches sometimes support the identification, clustering, classification, recognition works along with dimensionality that are necessary in terms of developing biometric systems[40]. Working on ***(ML)*** has been critically accelerated in the current years. Machine learning procedures have made the functioning of this biometrics recognition and identification more possible including that has also created much more advancement in this biometric method of recognition [3]. This "Biometric technology" can be considered as a real timing system that generally verifies a human's identity through utilizing their different characteristics basically as fingerprint patterns including facial geometrics[13].

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

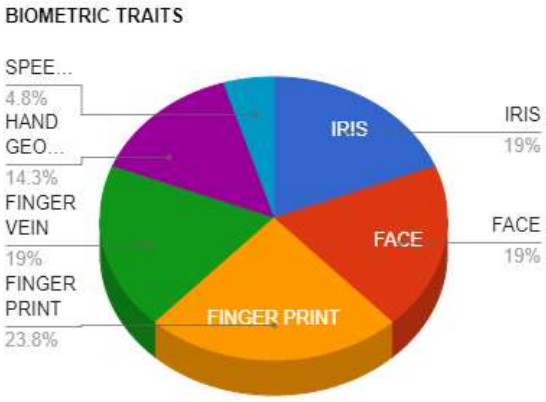


Fig. 3. The percentage of distinct biometric traits employed is depicted.

Machine learning utilizes a massive estimation of structured data in case that it might be able to produce outcomes or else provide several predictions depending upon those provided data [12]. In the present scenario, this ML has been utilized in identification fingerprint recognition, finger vein and facial recognition procedures. These machine learning oriented approaches generally aid in the identification, classification along with recognition-oriented functions that are needed in terms of developing these biometric-systems[4][39].

SL has been supporting numerous biometric applications by a massive estimation of algorithms. Few from those algorithms are Kernel Methods, Decision trees, Convolutional neural oriented nets including that logistic regression. The Decision-Tree is a well-known SL algorithm in which a decision is made at each node [29].This facial related recognition utilizes the "Decision-trees" oriented algorithm regarding the purpose of the authentic recognition biometrics system[14]. Moreover, several new surveys have stated that this naturally provides 100% accuracy including authenticity upon the "***Facial Recognition Technology Database " (FERET)***. Speech emotion-oriented recognition utilizes "***Support Vector Machines(SVM)***" algorithm including the '***Kernel Method***' named approach is utilized regarding facial emotion related recognition [14]. The SVM algorithm which is a SL algorithm is mostly used to evaluate classification of data [30]. Unsupervised machine-oriented learning has been implemented in these biometrics systems and it ensures effective learning-oriented methods, providing allowance regarding great classification including the authentic positioning in these biometric features [9]. This ***USL procedure*** sometimes can be utilized regarding extracting the overall automatic finger vein related pattern. Moreover, it can naturally be utilized in the initial stage for the purpose of effective learning strategy-oriented definition, feature fusion including data analysis and others[1].

The ultimate approach in this machine learning is ***Reinforcement ML(RML)*** can be considered as more versatile than the other supervised including unsupervised learning related approaches [7]. This RML approach is recognized as a strong ML approach that is methodically understood in performing several tasks along with training several algorithms through utilizing a method of reward including punishment. There are several software agents that naturally receive some kind of rewards and value after executing a function meticulously and penalties regarding an improper intervention[1]. This ML approach concentrates upon an action that is followed through an enhanced performance or those actions that is not into confederation of practice again. As per the study, it can be concluded that the fusion of these ML oriented algorithms including biometrics has given effective biometric system related accuracy including authenticity in comparison to the other conventional methods.

ML naturally helps in studying biometric topographies in terms of stimulating a separate person's identification of several learning activities[2]. In addition, all these approaches support the process of identification, clustering, recognition, classification including that recognition focused tasks that are extremely necessary in order to develop biometric systems. ML also has created the biometrics functioning identification more possible including that has also created much more advancement in the pattern of biometric recognition. USL might be utilized regarding total automatic finger vein related pattern extraction. This process of ML might be considered to be an outstanding approach in terms of obtaining the actual pattern of biometric recognition. This USL is basically utilized in voice detecting, matching of retina including that it also vastly utilized in fingerprint recognition. SL can be known as a modern technology about learning a particular function that mainly infers a proper role from the labeled training related data that outline an output depending upon sample of input-output pairing [12]. Some algorithms of this SL are ***Convolutional Neural Nets (CNN)*** [17]***, logistic regression, Kernel methods*** basically as ***SVM and Kernel Perceptron*** including ***Decision Trees***.

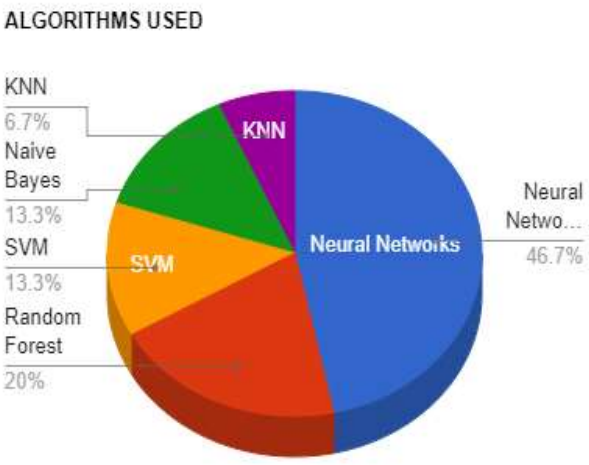


Fig. 4. Depicts the algorithms employed by various authors for biometric recognition from 2017 through 2022.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

IV. CONCLUSION AND FUTURE WORK

After reviewing various research papers on unimodal and hybrid recognition technologies, we found that most authors rely on ML technology for this purpose. More specifically, many of them chose neural networks (NN) for the proposed system. Also, the accuracy achieved by variations of the NN algorithm was generally over 95%. Therefore, the NN algorithm is one of the best algorithms for this purpose. We also learned that fingerprint, face, and iris biometrics are more reliable for identification purposes. We faced a variety of challenges, but the most difficult was determining the right algorithms and human characteristics. ML can be considered as a method or other procedure of this data analysis that naturally automates in making analytical models. “Hybrid Biometric Platform” can be recognized as a multimodal biometrics procedure that helps in making identification of finger vein, face recognition, and fingerprint identification[3][4]. Machine learning can also support in studying biometric topography in order to stimulate a prepared identification learning and gaining knowledge-oriented activities. Biometrics can be recognized as the technique that might be utilized in terms of identifying a separate person. As an example, fingerprint-oriented recognition can be utilized in order to identify some public through each other with the support of utilizing those people’s fingerprints. The Hybrid Biometric Recognition focused approach is mostly important for identification along with verification in single or else numerous systems.In the future, we'll aim to combine more biometric features to improve the HBRS and raise the accuracy of human recognition.

REFERENCES

[1] Carleo, G., Cirac, I., Cranmer, K., Daudet, L., Schuld, M., Tishby, N., Vogt-Maranto, L. and Zdeborová, L., 2019. Machine learning and the physical sciences. Reviews of Modern Physics, 91(4), p.045002.

[2] Moosavi, S.M., Jablonka, K.M. and Smit, B., 2020. The role of machine learning in the understanding and design of materials. Journal of the American Chemical Society, 142(48), pp.20273-20287.

[3] Prabu, S., Lakshmanan, M. and Mohammed, V.N., 2019. A multimodal authentication for biometric recognition system using intelligent hybrid fusion techniques. Journal of medical systems, 43(8), pp.1-9.

[4] Khan, Amreen, and Abhishek Bhatt. "Hybrid Biometric Recognition using Stacked Auto Encoder with Random Forest Classifier." *aging* 4: 5.

[5] Hossain, M.A. and Al Hasan, M.A., 2020. Improving cloud data security through hybrid verification technique based on biometrics and encryption system. International Journal of Computers and Applications, pp.1-10.

[6] Mohsin, A.H., Zaidan, A.A., Zaidan, B.B., Albahri, O.S., Albahri, A.S., Alsalem, M.A. and Mohammed, K.I., 2019. Based blockchain-PSO-AES techniques in finger vein biometrics: A novel verification secure framework for patient authentication. Computer Standards & Interfaces, 66, p.103343.

[7] Kumar, R.H., Haripriya, D. and Murikipudi, S., 2018. Integrated Biometric Attendance and Informative System using Cloud. *International Journal of Pure and Applied Mathematics*, *118*(20), pp.2807-2819.

[8] Mudliar, P., 2020, April. Whither Humane-Computer Interaction? Adult and Child Value Conflicts in the Biometric Fingerprinting for Food. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1-12).

[9] Bora, A., Vajpai, J. and Gaur, S., 2018. Speaker identification for biometric access control using hybrid features. *International Journal of Computational Science and Engineering, vol*, *9*, p.04.

[10] Oloyede, M.O., Hancke, G.P. and Myburgh, H.C., 2018. Improving face recognition systems using a new image enhancement technique, hybrid features and the convolutional neural network. *Ieee Access*, *6*, pp.75181-75191.

[11] Muliono, Y., Ham, H. and Darmawan, D., 2018. Keystroke dynamic classification using machine learning for password authorization. *Procedia Computer Science*, *135*, pp.564-569.

[12] Ahmadi, N. and Akbarizadeh, G., 2017. Hybrid robust iris recognition approach using iris image pre-processing, two-dimensional gabor features and multi-layer perceptron neural network/PSO. *Iet Biometrics*, *7*(2), pp.153-162.

[13] R. Chu, S. Liao, Y. Han, Z. Sun, S. Z. Li and T. Tan, "Fusion of Face and Palmprint for Personal Identification Based on Ordinal Features," 2007 IEEE Conference on Computer Vision and Pattern Recognition, 2007, pp. 1-2, doi: 10.1109/CVPR.2007.383522.

[14] Rizqyawan, M. Ilham, et al. "Comparing Performance of Supervised Learning Classifiers by Tuning the Hyperparameter on Face Recognition." International Journal of Intelligent Systems & Applications 13.5 (2021).

[15] Mohmand, Abdullah, et al. "Hybrid Model for Gait Biometric Recognition Using Machine Learning." (2020).

[16] Aly, Ola M., et al. "Multimodal biometric system using iris, palmprint and finger-knuckle." *International journal of computer applications* 57.16 (2012).

[17] Das, Rig, et al. "Convolutional neural network for finger-vein-based biometric identification." *IEEE Transactions on Information Forensics and Security* 14.2 (2018): 360-373.

[18] Zhang, Liping, et al. "ARFace: attention-aware and regularization for face recognition with reinforcement learning." *IEEE Transactions on Biometrics, Behavior, and Identity Science* (2021).

[19] Jain, Anil K., Arun Ross, and Salil Prabhakar. "An introduction to biometric recognition." *IEEE Transactions on circuits and systems for video technology* 14.1 (2004): 4-20.

[20] Suzuki, Hiroyuki, et al. "Secure biometric image sensor and authentication scheme based on compressed sensing." *Applied optics* 52.33 (2013): 8161-8168.

[21] Kumar, Ajay, and David Zhang. "Hand-geometry recognition using entropy-based discretization." *IEEE Transactions on information forensics and security* 2.2 (2007): 181-187.

[22] Patro, Kiran Kumar, and P. Rajesh Kumar. "Machine learning classification approaches for biometric recognition system using ECG signals." *Journal of Engineering Science & Technology Review* 10.6 (2017).

[23] Li, Haoxiang, et al. "A convolutional neural network cascade for face detection." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.

[24] Mohsin, A. H., et al. "Real-time remote health monitoring systems using body sensor information and finger vein biometric verification: A multi-layer systematic review." *Journal of medical systems* 42.12 (2018): 1-36.

[25] Saboor, Abdul, et al. "Latest research trends in gait analysis using wearable sensors and machine learning: A systematic review." *Ieee Access* 8 (2020): 167830-167864.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

[26] Nappi, Michele, Stefano Ricciardi, and Massimo Tistarelli. "Context awareness in biometric systems and methods: State of the art and future scenarios." *Image and Vision Computing* 76 (2018): 27-37.

[27] Jacob, I. Jeena. "Capsule network based biometric recognition system." *Journal of Artificial Intelligence* 1.02 (2019): 83-94.

[28] Liu, Ming, et al. "Fuzzified image enhancement for deep learning in iris recognition." *IEEE Transactions on Fuzzy Systems* 28.1 (2019): 92-99.

[29] Rahul and R. Katarya, "A Review: Predicting the Performance of Students Using Machine learning Classification Techniques," 2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2019, pp. 36-41, doi: 10.1109/I-SMAC47947.2019.9032493.

[30] Rahul and R. Katarya, "Impact of Supervised Classification Techniques for the Prediction of Student's Performance," 2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC), 2020, pp. 537-543, doi: 10.1109/I-SMAC49090.2020.9243360.

[31] Mady, Huda, and Shadi MS Hilles. "Face recognition and detection using Random forest and combination of LBP and HOG features." *2018 international conference on smart computing and electronic enterprise (ICSCEE)*. IEEE, 2018.

[32] Rahul, Ayush, D. Agarwal and D. Vijay, "Genre Classification using Character Networks," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 216-222, doi: 10.1109/ICICCS51141.2021.9432303.

[33] Ponnusamy, Vasaki. "Hybrid Behavior-based Biometric Authentication Systems (HBAS)." *Turkish Journal of Computer and Mathematics Education (TURCOMAT)* 12.10 (2021): 3013-3023.

[34] Rahul, V. Gupta, V. Sehra and Y. R. Vardhan, "Ensemble Based Hinglish Hate Speech Detection," 2021 5th International Conference on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 1800-1806, doi: 10.1109/ICICCS51141.2021.9432352.

[35] Raju, K., et al. "An Optimal Hybrid Solution to Local and Global Facial Recognition Through Machine Learning." *A Fusion of Artificial Intelligence and Internet of Things for Emerging Cyber Systems*. Springer, Cham, 2022. 203-226.

[36] Shaheed, K., Qureshi, I. A hybrid proposed image quality assessment and enhancement framework for finger vein recognition. *Multimed Tools Appl* (2022). https://doi.org/10.1007/s11042-021-11877-x

[37] Shakil, Samiya, Deepak Arora, and Taskeen Zaidi. "Feature based classification of voice based biometric data through Machine learning algorithm." *Materials Today: Proceedings* 51 (2022): 240-247.

[38] Winston, J. Jenkin, et al. "Hybrid deep convolutional neural models for iris image recognition." *Multimedia Tools and Applications* (2021): 1-23.

[39] Mugunthan, S. R., and T. Vijayakumar. "Design of improved version of sigmoidal function with biases for classification task in ELM domain." *Journal of Soft Computing Paradigm (JSCP)* 3.02 (2021): 70-82.

[40] Kumar, T. Senthil. "Video based traffic forecasting using convolution neural network model and transfer learning techniques." *Journal of Innovative Image Processing (JIIP)* 2.03 (2020): 128-134.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:06:03 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760968

A Deep Learning Approach for Interior Designing of Apartment Building Architecture using U2 Net

Gaurav Kumar
Dept. of Software Engineering,
Delhi Technological University,
Shahbad Daulatpur Village,
Rohini, Delhi, 110042.
gauravkumar_2k18se060@dtu.ac.in

Ashish Kumar
Dept. of Software Engineering,
Delhi Technological University,
Shahbad Daulatpur Village,
Rohini, Delhi, 110042.
ashishkumar_2k18se041@dtu.ac.in

Rahul
Dept. of Software Engineering,
Delhi Technological University,
Shahbad Daulatpur Village,
Rohini, Delhi, 110042.
rahul@dtu.ac.in

Abstract – The introduction of Deep Learning in computer technology has mastered various fields of study like medical, Banking, Finance etc. but the field of construction is lagging behind by still requiring some manual work. Apartment building designing is a complex task which requires professional architects and various tools. It would be easier and more fascinating if one can get the interior design just by having the shape of the land area. By considering this, a model has been proposed to serve the purpose of designing apartment architecture using generative adversarial networks. This study describes a system consisting of two sequential U²Net [13] based GAN models in which the first model has been trained for printing footprint of a building whereas second one for showing segmented sections for various areas of apartment.

Index Terms **- Deep Learning, Architectural Engineering, GAN, U²Net**

I. INTRODUCTION

Apartment building floorplan generation is a complex-manifold process and generally done by professional architects who may have years of experience in designing floorplans for buildings. Although we can hire an architect, this process becomes costly and time consuming. First of all being costly it's not affordable for everyone and poor designing of houses may result in inefficient use of land area. So these problems present a serious demand for having a system which can design the floorplan of our building very cheaply and in no time. Thus to save money and time we propose a novel U²NetGAN architecture for Apartment Design Model which uses land area shape to produce one of the efficient designs as shown in Fig. 1. Our model first uses image of land area as an input which is passed through our model. In this we have to identify various areas and edges, so a better feature detection technique was required, for that we have used U²Net model which performs very well in object detection scenarios.

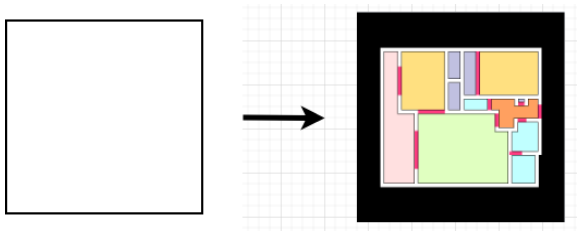


Fig. 1: Land Area to Segmented Floor plan

Considering this fact we have incorporated this architecture for the implementation of generator of our U²NetFloorPlanGAN model. After first pass, we will get footprint for the corresponding land area, now this land area will be passed to the next similar U²NetFloorPlanGAN architecture that has been trained for generation of segmented colored region for various areas of apartment using footprint of the apartment.

GANs are a method or an approach used in generative modeling. Generative modeling is an unsupervised process of learning models which consists of autonomously detecting and learning some patterns or figures in an input data so that the model can generate or produce new models that probably could be implemented from original data [2]. The aim of this study represents application of U²NetGAN in the field where they are not in much use like apartment designing. The model requires two datasets for the training of two sequential models. After training, the model has been tested using test samples. Performance of the model has been judged quantitatively.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

II. RELATED WORK

Although GANs have a wide research history but architectural engineering does not have much wide scope research behind them, the most significant work in this field has been done by Q Chen et. al. [5] in which they were creating 3D home design model using linguistic descriptions that is provided by the users by using NLP and GC LPN. They are first converting language input into structural graph representation which is then used to generate interior design of the house and outline of the rooms. They have also created text to 3D home model dataset for their study. Apart from this Wexin Huang et. al.[1] has done architectural designing using pix2pixHD during their study which is a modified version of GAN network. They have used it in recognizing and displaying the room(s) in several colors, after that it making floorplans with the help of CNN. Siu-Hang Or et.al.[6] have designed an approach to create 3D model of a house by understanding the designs of 2D floorplan images. They have created a model that took floorplan as a vector image and then changed it to vector outlines. Their model is generating interior and exterior design of a house. Morteza Rahbar et. al. [7] designed a model for generating architectural 2D designs on a particular footprint using agent-based designing with conditional GANs. Ahn et. al. [8] presented a novel architecture research method for floorplan images using deep learning techniques. They have also created their own dataset of apartments and then analyzed it through their algorithms.

In our novel proposed work, we are creating apartment designing model using generative adversarial networks. In previous study Stanislas Chaillou [11] has also used GANs for floorplan generation but they have incorporated pix2pix for their model. The results they got doesn't have very clear edges and sections were overlapping, so we have proposed this model in which edges will be sharp and overlapping of regions will be reduced. For this, we have used U2Net generator which learns mapping of different regions very well and for having sharp edges we have used a separate edge-loss function.

III. OBJECTIVE

We have proposed an "Apartment Designing model" using U^{2}NetGAN. The primary objective behind our research is specific to architecture field to develop GAN model for generating apartment design architecture. Our main motive behind this research is to explore architectural field with the help of AI in which not much research has been done over the years. We have used U^{2}Net [13] architecture for salient detection of apartment sections. This digital model can serve the purpose of architecture designing very cheaply and in future there may be many software programs available for doing such tasks even for free.

IV. ABOUT DATASET

This study requires two datasets for the model training: 1. Dataset for footprint design training (Land Area mapped to Footprints) 2. Dataset for interior design using footprints (Footprints mapped to Corresponding Interior Design).

A. Preparation and Preprocessing of first Dataset

We have used the Microsoft repository of USBuildingFootprints [3] for the preparation of the first dataset. It contains the building shape of apartments of a society from the top as shown in Fig. 2.

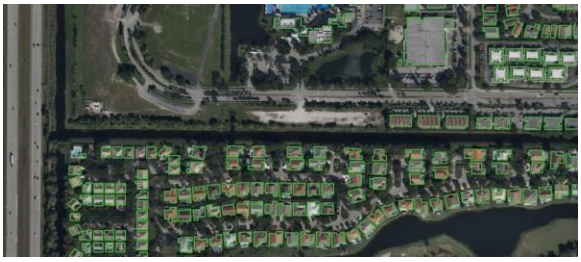


Fig. 2: US Building Footprints [3]

We have written scripts to extract data from geographical coordinates of US buildings footprints datasets, then these footprints were superimposed on a white background. For each image, a bounding box was created using the '*lableme'* tool. Our next step is to map the land area images to corresponding apartment building images.

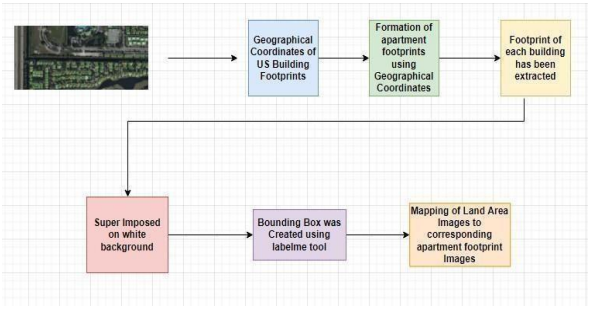


Fig. 3: Dataset Generation for Model 1

Now we need to label the images for this purpose, *'LabelImg'* software has been used. So we have generated a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

total of 505 input-output images of dimension (256 × 256) for training and 130 input- output images for testing. All the steps for the preparation of the first dataset have been shown in Fig. 3. Snapshots of the generated dataset for the first stage of the model have been shown in Fig. 4 & Fig. 5.



Fig. 4: Input Images for the first dataset

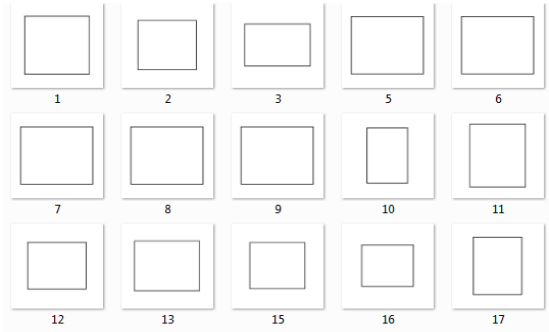


Fig. 5: Output Image for the first dataset

B. Preparation and Preprocessing of Second Dataset

For the preparation of the second dataset, we have used the dataset "r2v " floor_plan from the zlzeng github repository named DeepFloorplan [4]. This dataset contains 679 sets of train data images where each set consists of 5 images (close, close_wall, multi, rooms, wall) whereas test data consists of 100 sets where each set has 3 images (close, close_wall, rooms). Following steps were followed for preparation of dataset:-

1. We have taken 'rooms' images from train and test data to prepare output data for second data.
2. We generated input data for the second dataset from 'rooms'.images for both train and test data.
   - We detected the contours to get external points of edges of footprint images.
   - After detecting the contours we found corner points for the footprints from the set of contours.
   - After creating corners we then created one black image of the same shape as that of footprint image and filled white colors inside geometry formed by the detected corner.

Our generated dataset consists of 679 training input- output images and 100 testing input-output images of dimension (256× 256). We can see the snapshot of the dataset for the second model in Fig. 6 & Fig. 7.

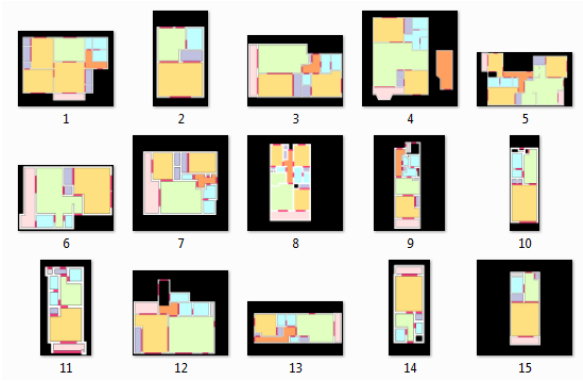


Fig. 6: Input Image for the second dataset

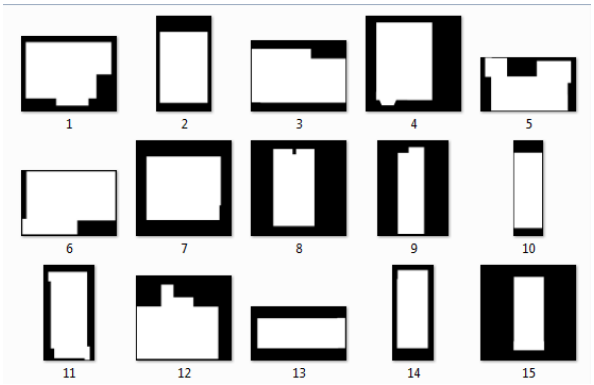


Fig. 7: Output Image for the second dataset

V. METHODOLOGY

A. Proposed Model

The proposed Apartment Design Model is a floor plan designing model which consists of 2 steps where both steps contain a GAN. Here, we sum up how the model designs the entire structure of the building. First of all we normalized the input image to (256 × 256 × 3) before passing to the first generator network.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

Step 1: Building Footprint

Building footprints extensively describes the inner structure of floor plans. Their outline is dependent on the shape of the building. Land shape may be used to deduce the design and positioning of footprint of house, we train our model for generating footprints through our dataset.

Step 2: Creating Segmented Regions

Step 2 handles segmentation of regions. Our model takes the building footprint of the house which was generated in step 1 as an input, the location of the entrance, rooms, living area, bathroom & main windows defined from the user. The floor plans images are utilized to train our model. In the output, the program describes each room by using different colors.

Below Fig. 8 shows each step of the workflow (I) is Building footprint & (II) is Creating Segmented Regions.

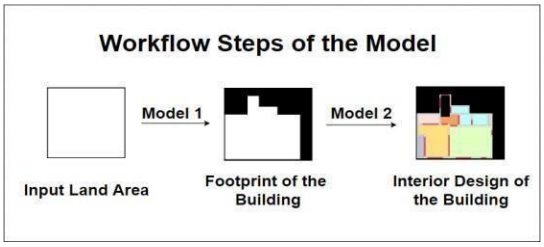


Fig. 8: Workflow steps of the model

In both stages we have used a GAN model with the help of U^{2}Net [13] architecture in the generator network which has been trained with their corresponding dataset. The building footprint images that will be generated in step 1 are processed in the next part of our model to create segmented interior design to have different locations in the apartment.

B. Generative Adversarial Network

Generative Adversarial Networks (GANs) are deep learning models based on the concept of generative modeling. They use a supervised learning approach in which the goal is to learn mapping of input to output. The GAN model has two sub models: Generator and Discriminator [2]. The job of the generator is to convert an input image to output image and it will generate it's converted image according to its learning at each epoch and this generated image will be passed to the discriminator network which is a classifier network whose job is to predict whether the image is original or has been generated by the Generator. Generator's aim is to trick the Discriminator network whereas Discriminator's aim is to keep distinguishing real output image from the image generated by the generator and keep analyzing the generator when it classifies the image as fake. This process keeps continuing until the Discriminator fails to differentiate between the original image and the image generated by the generator & so classify it as a true image. This competition between two parts of the model results in better quality of output image which closely resembles real image [2]. Fig. 9 shows the experiment design of GAN.

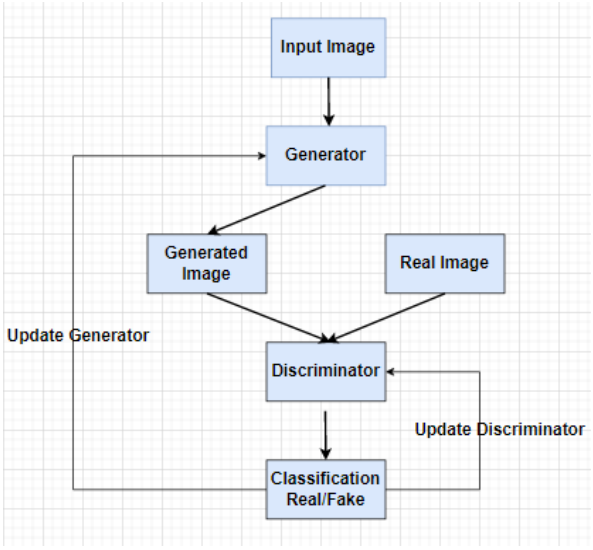


Fig. 9: Experimental Design for GAN

C. Generator Network

The generator network is responsible for creating new images by learning mapping between input and output images. We have passed (256 × 256 × 3) as input in the generator network. This architecture is based on U^{2}Net[13] architecture. The only difference is that unlike concatenating outputs of all networks we are creating just a single output using all of them.

D. Discriminator Network

The discriminator network is a classifier network whose task is to classify the generated image as real or fake. In this model two images; Image generated by generator and the target image both of (256 × 256 × 3) are passed as an input.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

Then these two are concatenated together to form (256 × 256 × 6) and then after few layers of down sampling, zero padding, batch normalization and leakyRelu as an activation function are applied sequentially in the subsequent layers finally we get an image of (30 × 30 × 1). Architecture for Discriminator has been shown in Fig. 10.

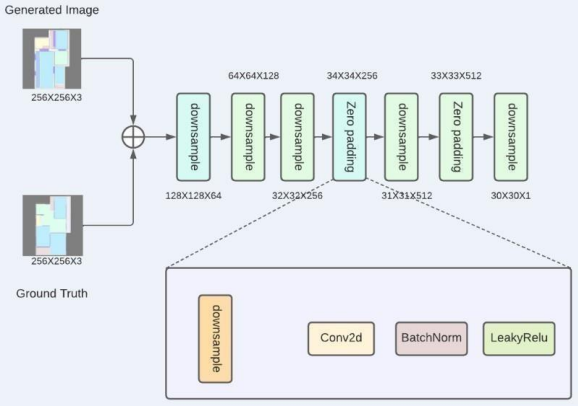


Fig. 10: Discriminator Network

E. Model Training

When discriminator classifies image as fake/real then losses for both discriminator and generator are calculated on the basis of discriminator's output. Then it needs to backpropagate to obtain gradients and these gradients then result in change in weights in the right direction. The training of both sub models is done alternatively (parallel training). When the discriminator training phase is running then the generator will remain constant and when the generator is being trained the discriminator will keep constant. After training each time the generator will create a better image then the previous one but now the discriminator has also been trained so its classification has also become better. So this parallel training results in generation of better image after epoch.

F. Edge Loss Function

Edge Loss function has been used for sharp edges of the apartment. Edge loss function uses Sobel Edge Detector for the calculation of loss. This function first creates edges from the output and ground truth image and then calculates edge loss using Eq. 1. Using this loss edges keeps getting better at each epoch.

$$L_E = 1/N \left(\sum \left(d_{ix} - d_{ix}^{GT} + d_{iy} - d_{iy}^{GT} \right) \right) \quad (1)$$

where,

$$dix = di \times \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \quad (2)$$

$$diy = di \times \begin{bmatrix} -1 & -2 & -1 \\ 0 & 0 & 0 \\ 1 & 2 & 1 \end{bmatrix} \quad (3)$$

Where, **N** is the total number of pixels in the final output depth map, **dix** and **diy** are sobel gradients in **x** and **y** direction respectively and GT is groundtruth, as shown Eq. 2 and Eq. 3.

VI. RESULTS AND FINDINGS

A. Qualitative analysis

We have got significant results in this study and the designs of apartments generated on the test dataset were pretty satisfactory. Our model generates footprints of different shapes based on different boundaries image passed to the model and generated output from model 1 were almost similar to the groundtruth footprint as shown in Fig. 11.

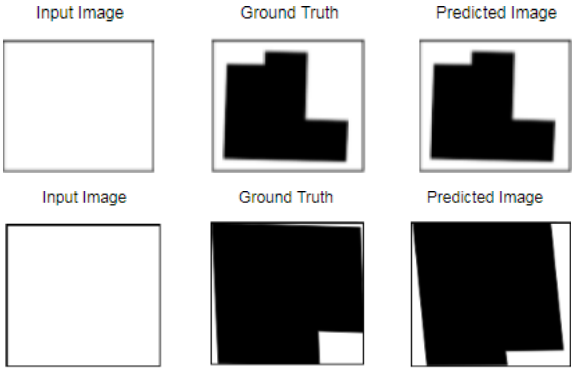


Fig. 11: Input, Ground Truth and Predicted Image from Stage 1

Stage 2 model is capable of generating different apartment designs on the basis of different footprint passed, it is successful in defining edges and different regions very well and qualitatively results are pretty good as we can see the results in Fig. 12.

From visual analysis we can say that both stages of Apartment Design Model performed pretty well and few designs were using the space of land area even better than ground truth.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

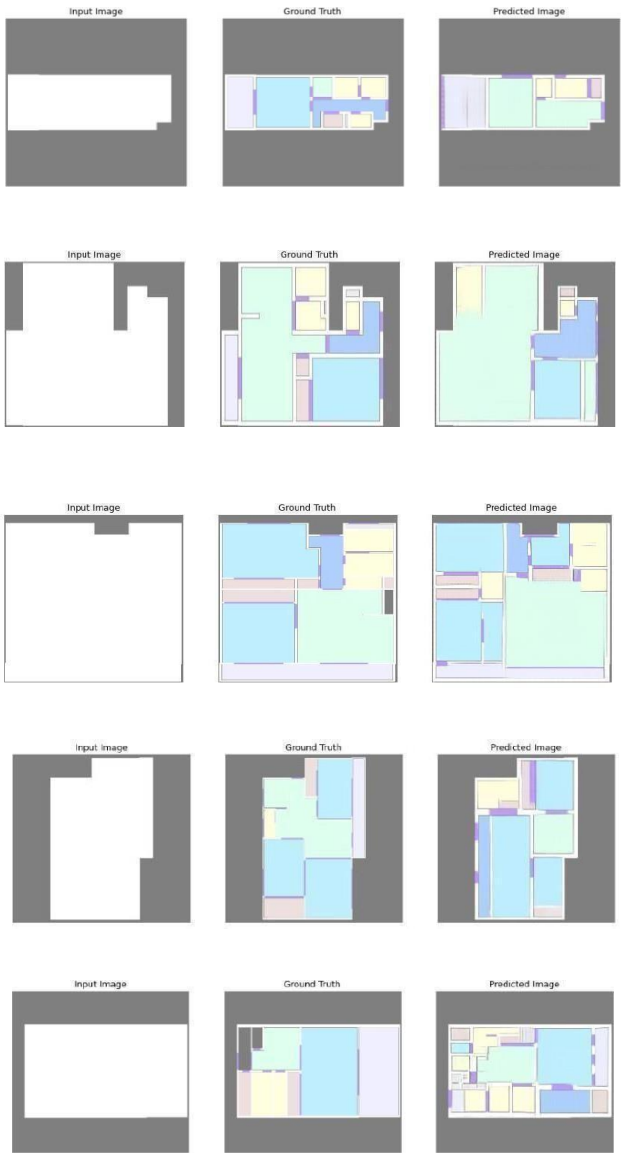


Fig. 12: Various Input, Ground Truth, and Predicted Images from Stage 2

B. *Survey Analysis*

Apart from analyzing qualitatively, we have also used public survey method for predicting our score of how accurately our model predicted the output. For this, we have a created a Google form in which we have asked to rate quality of predicted image as per the groundtruth image on a scale of 1-5 with 1 being poor and 5 being excellent. We have got 250 responses and results of those responses are shown in the given Fig. 13.

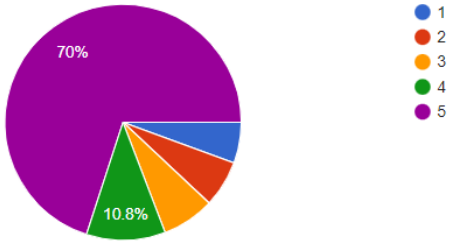


Fig. 13: Response of public survey

We have calculated the expected average value using these responses and it comes out to be 4.33 out of 5 i.e. 86.6% that indicates our predicted images are quite similar to groundtruth images and that means our model generate good results.

VII. CONCLUSION

In this study we have proposed our U2NetFloorPlanGAN model which has been used in floorplan generation. So in this study we have shown how we took GAN technology to another level. We conclude that we can make use of GAN in any field. One of the main advantages of our model is that it classifies the regions very well with the help of U2Net model as it learns the mapping of various regions according to its place and it gives appropriate places to different sections of apartment. Overlapping extent of various regions has been minimized to nearly zero as they are separated by sharp edges that conclude that our model is generating predicted image very accurate with respect to groundtruth image. Based on our public survey and the results that we got from it is quite impressive, we conclude that our model performed really well. With the help of this study we tried to expand the scope of GANs by exploring the field which is lagging behind in making use of DL i.e. the architectural field. On taking a close look at the generated outputs we can say that GAN technology will definitely adapt in the field of construction very well.

VIII. FUTURE WORK

We have tried and tested our model only on rectangular shapes and this model is capable of working well on straight edges. In future we would like to extend this for curve edges also by adding more loss functions and hybrid dataset containing images of both straight and curve edges. Whenever we call an architect he/she simply asks for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 UTC from IEEE Xplore. Restrictions apply.

various information about designing like number of rooms, width of stairs etc. and after gathering this information he/she gives us a single design for the apartment. Even after paying such a huge amount we get only a single design without having any other option. Hence, we will try to solve this problem and introduce variability factors in predicted images, so that our model will be able to generate multiple designs and we will have more options to choose from.

IX. ACKNOWLEDGMENT

We would like to thank Microsoft's public github repository "US Building Footprints" [3] and zlzeng's github repository "DeepFloorplan"[4] as the source images for creation of our dataset. We are also thankful to Qin et. al. for introducing the U^{2}Net[13] architecture which we have used in our model.

We would also like to thank Mr. Rahul for his active guidance throughout the research. All the research work implementation and research paper writing was carried out under his constant assistance.

X. REFERENCES

[1]. Hao Zheng, Weixin Huang, "Architectural Drawings Recognition and Generation through Machine Learning", Conference: Proceedings of the 38th Annual Conference of the Association for Computer Aided Design in Architecture (ACADIA), ISBN 978-0-692-17729-7] Mexico City, Mexico 18-20 October, 2018, pp. 156-165.

[2]. Jason Brownlee. 2019, "A Gentle introduction to Generative Adversarial Networks (GANs)". [Online]. Available: https://machinelearningmastery.com/what-are-generative-adversarial-networks-gans/ [Accessed: March 26, 2022].)

[3]."microsoft/USBuildingFootprints: Computer generated building footprints for the United States". [Online]. Available: https://github.com/microsoft/USBuildingFootprints. [Accessed: March 26, 2022].

[4]."zlzeng/DeepFloorplan." [Online]. Available: https://github.com/zlzeng/DeepFloorplan. [Accessed: March 26, 2022].

[5]. Q. Chen, Q. Wu, R. Tang, Y. Wang, S. Wang, and M. Tan, "Intelligent home 3D: Automatic 3D-house design from linguistic descriptions only," Feb. 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020, [Online]. Available: http://arxiv.org/abs/2003.00397.

[6]. Sui-Hang Or, K.H. Wong, Yinglin Yu and Michael Ming-Yuen Chang (2005) "Highly Automatic Approach to Architectural Floorplan Image Understanding & Model Generation", The Chinese University of Hong Kong, Shatin, N.T., Hong Kong.

[7]. Morteza Rahbar, Mohammadjavad Mahdavinejad, Amir H.D. Markazi, Mohammadreza Bemanian, "Architectural layout design through deep learning and agent-based modeling: A hybrid approach", Journal of Building Engineering, vol. 47, p. 103822, Apr. 2022, ISSN 2352-7102. doi: 10.1016/J.JOBE.2021.103822.

[8]. Ahn et.al. "An Analysis on the Evolution of Korean Apartment Unit Plans using Deep Learning," Journal of the Architectural Institute of Korea , vol. 37, no. 10, pp. 13–22, Oct. 2021, doi: 10.5659/JAIK.2021.37.10.13.

[9]. Yusuf Cihat Aydin; Parham A. Mirzaei; and Jonathan Hale, "Using machine learning techniques to predict esthetic features of buildings," Journal of Architectural Engineering, vol. 27, no. 3, p. 04021023, Sep. 2021, doi: 10.1061/(ASCE)AE.1943-5568.0000477.

[10]. Stanislava Fedorova. 2021. "Generative Adversarial Networks for Urban Block Design" in SimAUD 2021: ASymposium on Simulation for Architecture and Urban Design, April 15–17, 2021, Online. ACM, New York, NY, USA.

[11]. Stanislas Chaillou, "ArchiGAN: Artificial Intelligence x Architecture", In book: Architectural Intelligence, Selected Papers from the 1st International Conference on Computational Design and Robotic Fabrication (pp.117-127), 2020, doi: 10.1007/978-981-15-6568-7_8.

[12]. M. Bachl and D. C. Ferreira, "City-GAN: Learning architectural styles using a custom Conditional GAN architecture", July 2019, doi: 10.48550/arxiv.1907.05280.

[13]. Xuebin Qin, Zichen Zhang, Chenyang Huang, Masood Dehghan, Osmar R. Zaiane Martin Jagersand, "U2 -Net: Going Deeper with Nested U-Structure for Salient Object Detection", Volume 106, May 2020,107404, ISSN 0031-3203, doi: 10.1016/j.patcog.2020.107404.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:07:13 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 authentication scheme for group based communication for IoT oriented infrastructure in smart cities

Irfan Alam (irfanrmamu@gmail.com)
Delhi Technological University https://orcid.org/0000-0001-8596-907X
Manoj Kumar
Delhi Technological University

Research Article

Keywords: Smart City, Internet of Things (IoT), Grouped communication, authentication, Secret sharing, Scheme

Posted Date: April 27th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1571581/v1

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

A novel authentication scheme for group based communication for IoT oriented infrastructure in smart cities

Irfan Alam, Manoj Kumar
Department of Computer Science and Engineering
Delhi Technological University, New Delhi, India, 110056
Email: irfanrmamu@gmail.com

***Abstract*—As a result of the rapid developments in digital technologies, smart cities have been adorned with several digital gadgets based on the Internet of Things (IoT). The essential part of smart cities operations is IoT communications. IoT devices generate, process, and exchange a large amount of security-critical and privacy-sensitive data regularly, resulting in attractive targets for threats and attacks. Group-oriented communication plays a critical role in the IoT world, such as data gathering and area monitoring. The lower time complexity of Group-based authentication (GBA) makes these protocols very popular for efficient and secured communication. This paper uses a polynomial-based group authentication scheme and membership verification to ensure efficient and threat-free communication among IoT devices. Furthermore, a Bi-variate polynomial is used instead of the single variable polynomial to add the novelty of the work. Along with novelty, the protected nature of the bi-variate polynomial makes the proposed scheme very secure and reliable. Security analysis of the proposed work shows its efficiency over existing schemes.**

keywords: Smart City, Internet of Things (IoT), Grouped communication, authentication, Secret sharing, Scheme.

I. Introduction

Without a broadly agreed-upon definition, the smart city sector is still in the "I'll know it when I see it" phase. According to the Smart Cities Council, a smart city is one where digital technology is incorporated throughout all city functions [1]. Almost all smart cities sectors are using the Internet of Things (IoT) for an easy and efficient medium for Intra and inter functionalities. From Fig. 1, we can easily conclude that Smart cities are the super application domain of IoT. Fernandez-Anezan [2] performs an analysis of the various descriptions of a smart city. Luong et al. [3] analyze the economic pricing policies and their relationships in communication and data collecting for IoT. In contrast, Arasteh et al. [4] and [5] discuss the relationship between the smart city and IoT. In 2021, Chai et al. [6] explored the Big data, IoT and Cloud for smart cities .

From Industrial Automation to Health Care, IoT devices have become an integral part of almost all activities of smart cities. Security is needed at every user hierarchy level, such as secure booting, access control, update, and patch. To assure security, authentication is mandatory for all sectors of IoT-enabled smart cities [7]. Several research papers have recently been published that deal with traditional IoT authentication as [8], [9], [10] and [11]. Traditional authentication mechanisms are not scalable for densely deployed IoT networks containing millions of nodes expected to be operational. To assure security, authentication is mandatory for all sectors of IoT-enabled smart cities [7]. Several research papers have recently been published that deal with traditional IoT authentication such as [8], [9], [10] and [11]. Traditional authentication methods that follow the establishment of the session key are not scalable for densely deployed IoT networks, where millions of nodes are expected to be operational.

Since the uses of IoT devices are increasing day by day leads to the exponential growth of data and devices. Security is needed at every user hierarchy level, such as secure booting, access control, update, and patch. Due to the massive quantity of data interchange, the proliferation of IoT devices in smart cities has resulted in several performance concerns, including excessive latency and network congestion. The way of communication in Many applications such as supply chain management, smart energy grids, area monitoring, flood sensing, fire alarm, and many more are based on the Group-oriented communication [12]. Authentication in IoT systems, in particular, is critical for smart cities and must be accessible, secure, and quick. To authenticate members in group-based communication, [13] proposes Group based authentication (GBA). GBA is one of the authentication solutions, such as massive authentication requests to the authentication server and congestion control. GBA1 is for synchronous, and GBA2 is for asynchronous communication. In [14], authors compare the existing schemes and conclude that Communication complexity is reduced in GBA. Group-based authentication plays a vital role in the case of a large number of nodes. The Foundation of GBA is Secret Sharing Scheme (SSS) by Adi Shamir in 1969 [15]. In SSS, a secret key is distributed among the members of a group in the form of a unique shadow (token) such that in a later stage, a certain number of members can reconstruct the secret using their tokens without the involvement of an authentication sever. However, tokens are disposable after participating in secret regeneration in the case of GBA1. Similarly, an adversary can recollect the tokens from the network and later can perform attacks such as spoofing attack [16]. Chien et al. [16] have used the Elliptic curve discrete logarithmic (ECDLP) to solve the problem of GBA2. Furthermore, resembling the IoT environment makes

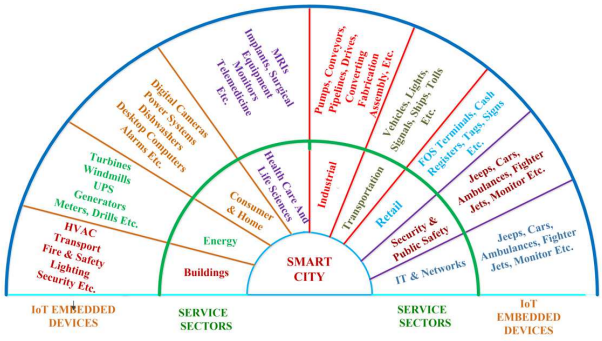


Fig. 1. Smart City and IoT World

chein's [16] scheme insecure [17]. However, Only as a pre-processing of user authentication can group authentication be used. because it cannot identify who non-members are.To identify non-members, additional one-to-one user authentications are required known as membership authentication [18], [19] and [20].

In Group based membership communication, One of the most critical aspects of the communication between the trusted authority and the participants in using private channels. A private channel multiplies the latency and is prone to several attacks such as man-in-the-middle attacks and spoofing attacks. Insider adversary is also one of the headaches in the attack-free implementation of group-based membership authentication. Our primary contributions are mentioned below in light of these challenges:

- We propose a scalable, lightweight authentication scheme that can be deployed in centralized or distributed group authentication scenarios to overcome possible pitfalls in group authentication.
- This scheme provides authentication among the members and generates a session key for uninterrupted communication.
- Due to resource-constrained nodes in IoT scenarios, lightweight schemes are mandatory, and We tried to keep the proposed scheme simple, flexible, and lightweight than the existing schemes.
- We have used bi-variate polynomial as in [21], [22] and [23], that makes our schemes very light as compared to others because it eliminates the requirement of extra key generation for the exchange of tokens among the nodes.

The rest of the article comprises related works in section II, followed by preliminaries and proposed protocol in sections III and IV. Security and performance analysis of proposed scheme is in section V. Section VI and VII contains conclusion and acknowledgment.

II. Related works

Authentication is critical for IoT devices in smart city infrastructure to avoid severe damages and mass destruction. Due to limitations of resources such as heterogeneity and low power supply, an authentication mechanism must be lightweight. Researchers have used many lightweight techniques such as ECC, blockchain,hash and XOR functions instead of heavy cryptographic algorithms [24], [25] and [26].

In the context of a smart city, several devices constitute a group. Secure group communication and membership authentication is needed for hours. In 2013, harn [13] proposed an idea of group authentication whose complexity $(O(n))$ is very much less than general approach of authentication $(O(n^2))$ [14]. Recently, [27], [28] and [29] have used group-based authentication in fog computing, drones network and Intelligent transportation system (ITS) respectively. In [16], Chien et al. demonstrates the flaws in Harn's asynchronous methods and propose a new scheme to overcome the issues such as spoofing attack and make it more efficient. To achieve his purposes, Chien et al. used the Elliptic Curve property known as Elliptic curve discrete logarithmic problem (ECDLP). In [17], lee tries to overcome the shortcomings in Chien [16] and Harn [13] schemes. In group-based authentication, an adversary can pretend to be a legitimate member, or a legitimate member may act as an internal adversary. To prevent such types of attacks membership authentication has been used by several authors such as [18], [19] and [30]. In the case of membership authentication, there is a need for a secure network to exchange tokens among other members to prevent attacks such as man in the middle attack and reply attacks. Setting up a secure network separately for each interchange of tokens/tokens will lead to inefficient and uneconomical Infrastructure of IoT devices that will never suit smart cities. In our scheme, we have proposed a novel scheme where one part of the token is used to encrypt another part.

III. Preliminaries

A. The (t, n) Threshold Scheme of Shamir

We have one secret in the Shamir Secret Sharing Scheme (SSS). Let's say it's C. C's shadows (tokens), designated as $F1, F2, ...Fn$, are allocated to n stakeholders (members). To regenerate the secret, a minimum k number of shadows are essential during the reconstruction phase. The scheme is known as the (t, n) threshold scheme. t is a threshold value in this scheme.
Mathematical concept behind SSS is Lagrange interpolation. Generally, To make the scheme more secure, Modular arithmetic is used in place of real-arithmetic, in secret sharing schemes. We can summarize the flow of SSS in following steps:

- A no.M is selected, which is prime and $M \geq C, n$.
- A function $g(x) = a_0 + a_1x + a_2x^2 + \cdots + a_{k-1}x^{k-1} \mod M$ is taken where $a_0 = C$ is the secret.
- Pair $(i, g(i))$ is generated by using $g(x)$, where $i = 1, 2, ...n$.
- These points $(i, g(i))$ are allocated among n participants/ members in secure manner.
- A minimum of t token holders collaborate in the reconstruction and exchange tokens through a secure channel.

$$g(x) = \sum_{i=1}^{n} L_i q_i \tag{1}$$

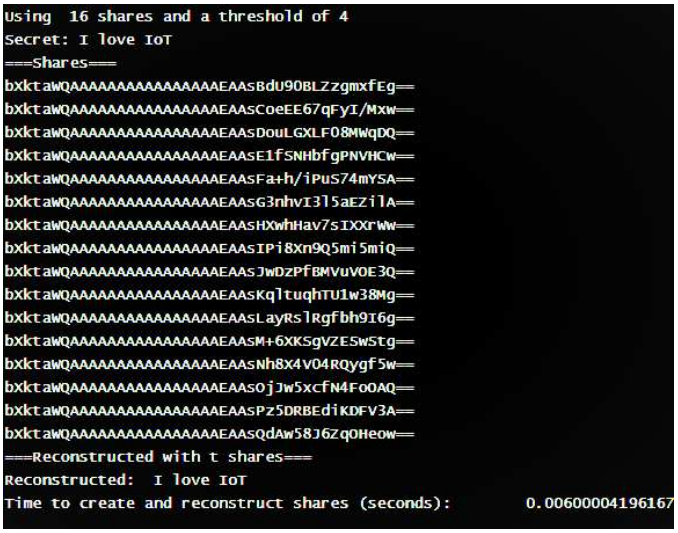


Fig. 2. Total time taken in secret sharing scheme

where $L_i = \prod_{j=1,j\neq i}^{n} \frac{x-x_j}{x_i-x_j}$. From (1), we get

$$g(x) = a_0 + a_1 x + a_2 x^2 + ... + a_{k-1} x^{k-1} mod Q$$

. Therefore $a_0 = C$ is the secret. The use of a secret sharing scheme in authentication would result in great savings of time. Time taken by general authentication methods for large no devices is much greater than the use of secret sharing. As in Fig 2, for the whole process of the secret sharing scheme time taken for 16 devices is 0.006 seconds. While in case of general authentication methods total time = $16 * 0.012 = 0.092 sec.$ as in [31].

Although Shamir's scheme is base of future works but there exist several shortcomings, some of them are following drawbacks of shamir secret sharing scheme.

- It works for single secret and unable to deal with multi-secrets.
- Recreation of token is mandatory once used.
- No communication between dealer and participants without private channel
- It is impotent to recognize the cheater or adversary.
- It aspires for the ideal situation, in which each token holder has the same priority or participation in the secret reconstruction phase.

There are several literature to overcome the drawbacks of shamir's secret sharing schemes such as [32], [33] and [34] for multi-stage and multiple secret sharing schemes. Similarly, [35] and others are used for verification of the participants token.

B. Group Based Authentication

Smart schools, smart water supply management, smart health treatment, smart waste management, autonomous transportation systems, smart grid system, and many other facilities are all part of the process of creating a smart city. Most of these systems work in a group, making grouped communication integral to smart cities' smooth and efficient setup.

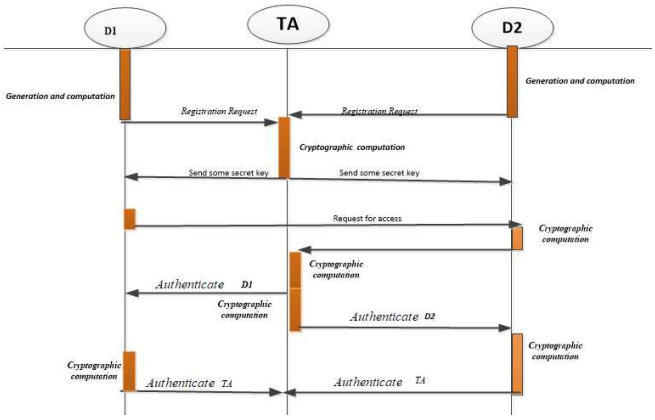


Fig. 3. Informal authentication process

Group based membership authentication is the fundamental security services in group based communications [36]–[40]. Authentication and identification management, in particular, are crucial and must be simple, reliable, and inexpensive. In a conventional authentication mechanism, there is one to one relationship of verifier and prover. From Fig.3, we may see a number of steps followed in conventional authentication, such as registration of IoT devices, generation of nonces, and verification techniques such as a digital signature. Each step carries cryptographic computation, which makes conventional authentication unfit for many low-powered IoT devices. On the other hand, grouped authentication consists of only three steps, i.e., distribution of tokens by a trusted authority, exchange of tokens among members of groups, and calculation of secret by each member on its own end. Grouped authentication steps can be seen in Fig.4. It is clear that group authentication needs less computational time and is thus very suitable for low-powered devices. Grouped authentication shows far better performance when no of nodes increases [14].

Group based authentication can be summarized briefly as follows :

- There is one Trusted authority (TA) sometimes called Group manager (GM) to distribution of tokens.
- To form a group, n group members register at the Trusted authority (TA)
- In registration phase , TA selects a random polynomial $g(x)$ with $g(0) = C$, with $(t-1)^{th}$ degree (where $t \leq n$) .
- TA computes secret tokens for members as $y_i = g(x_i)$.where xi denotes the public information for member Ui.
- Each member U_i, receives their token y_i from TA. The secret is released by the masking of one-way function $H(c)$.
- Each user releases the token obtained from the TA during the authentication phase. If all of the released tokens are legitimate, the polynomial $g(x)$ can be reconstructed and the secret obtained by interpolating the released tokens. Hashed form of this obtained value is compared to already public $H(C)$ value. If the result is same then all nodes authenticate each other, otherwise there is chance

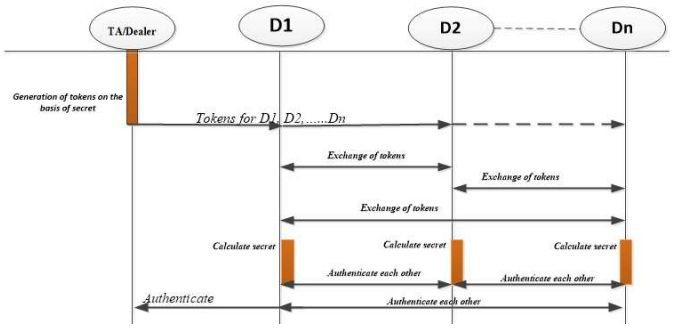


Fig. 4. group based membership authentication process

of adversary.

C. One-way Cryptographic Hash Function

A lightweight scheme designer must address the trade-off between security, costs, and performance. Mostly, authentication schemes use the hash function as one of the core parts to make their schemes efficient such as [41]–[43]. Detailed theories related to hash functions are available in [44]–[46]. Based on these detailed studies, we conclude the following properties of one–way hash function:

- This function is a deterministic function

$$H : \{0,1\}^* \rightarrow \{0,1\}^n$$

- $y = g(x)$ (called hash digest) is produced by hash function where output $y\epsilon\{0,1\}$
- Any minor modification to the input string can result in an entirely different hash value (message digest).
- Preimage-resistance —Finding any input that hashes to any pre-specified output is computationally impossible, i.e., finding any preimage x such that $g(x) = y$ when given any y for which a matching input is unknown.
- Collision resistance: finding any two separate inputs x, x' that hash to the same output, i.e. $g(x) = g(x')$, is computationally impossible.

D. System Model

The modular architecture of IoT based systems may vary application to applications. In the case of grouped communication, there is one trusted authority (TA), sometimes called the dealer. IoT nodes are represented as $\{U_1, U_2, ...U_n\}$ .Each user U_i interact with another user U_j.They exchange tokens and authenticate on one to many basis. Here in our scheme, there is no need of a secure channel.

E. Adversary model

There must be an adversary model against which the scheme is safe when used for secure authentication. Dolev presented an adversarial model for analyzing security protocols, and Yao [47] in 1983.

- All messages must go through the adversary.
- All messages can be read, altered, and redirected by an adversary. However, encryption works like a black box.

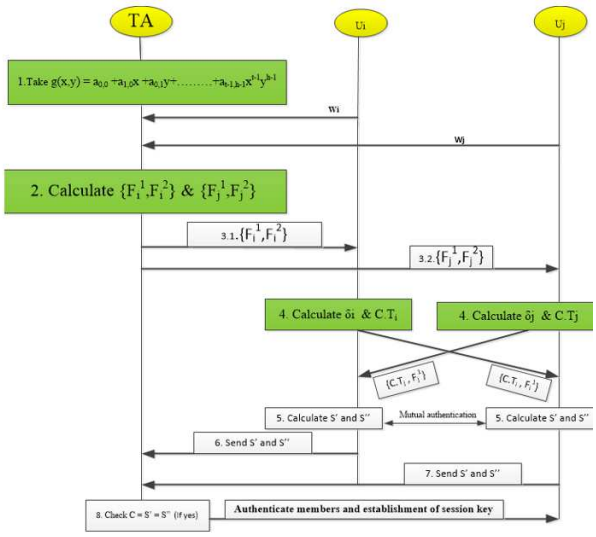


Fig. 5. group based membership authentication process

- If an adversary has the right keys, she can only perform decryption.
- She can only compose messages between new messages from keys and messages that she already possesses.
- Adversaries cannot perform any statistical or other crypt-analytic attacks

One thing that must be noted that adversary can be external and internal both.

IV. Proposed Scheme

A secure channel is needed for the exchange of tokens among the nodes as in [27], [28] and [29]which makes these schemes uneconomical with latency. In our work, we have proposed a novel group-based authentication scheme that eliminates secure channels' dependency. Also, members authenticate each other, and then a session key is established for future communication. The proposed scheme is based on a bivariate polynomial, and the Trusted authority chooses a bivariate polynomial. In our proposed protocol there is trusted authority (TA), n users $\{U_1, U_2, ...U_n\}$ and Authentication key is C. For particular group member U_i, TA calculate F_i^1, F_i^2.
In the authentication phase $H(C^{'})$ and $H(C^{''})$ are calculated, where H is one-way hash function. If the calculated value is the same, each member authenticates in the same group; hence the session key is established. The proposed scheme consists of two phases.

A. Pre-deployement phase

In this phase , TA perform following calculations:

- select a bi-variate symmetric polynomial

$$g(x,y) = a_{0.0}+a_{1,0}x+a_{0,1}y+\cdots+a_{t-1,h-1}x^{t-1}y^{h-1}\ (mod)p. \quad (2)$$

here $C = g(0,0)$ is the secret and $a_{i,j}\varepsilon Z_p$ where p is the large prime number.

TABLE I
NOTATIONS

Symbol	Explanation
TA	Trusted authority
U_i and U_j	$i^{th} and j^{th}$ users
w_i and w_j	public information of i^{th} and j^{th} users
δ	lagrange component of i^{th} user
P	large prime number
$H()$	Homomorphic function
$\sum$	Summation
$\prod$	Product
$\oplus$	XOR operation

B. Registration phase

- Each member of the group $\{U_1, U_2, ...U_n\}$send its publicly known values i.e,$\{w_1, w_2.....w_n\}$
- TA calculates pair of shadow $\left\{F_i^1, F_i^2\right\}$ for every member U_i in which $F_i^1 = g(w_i, 0) and F_i^2 = g(x, w_i)$.

C. Authentication phase

In authentication phase, each members performs following calculations:

- In reconstruction phase ,there are u no.of tokens holders participates where , $(t \leq U \leq n)$.
- Using shadow's second part, i.e., F_i^2 each member computes Lagrange's component

$$\delta_i = F_i^2(0) \prod_{j=1, j\neq i}^{u} \frac{-w_i}{w_i - w_j} \ (mod)\mathbb{P}. \tag{3}$$

- member U_i and U_j calculate $C.T_i = \delta_i \oplus F_i^1$ and $C.T_j = \delta_j \oplus F_j^1$ respectively.
- Each member send pair $(C.T_i, F_i^1)$ to other members separately.
- After getting pairs from all other members. Each member computes

$$S^{'} = \sum_{j=1}^{k} (\delta_j) \tag{4}$$

. and

$$S^{''} = \sum_{j=1}^{n} (F_j^1) \prod_{j=1, j\neq i}^{u} \frac{-w_i}{w_i - w_j} \ (mod)\mathbb{P}. \tag{5}$$

- If the above the calculated result from equation (4) and (5) are same i.e $H(S^{'}) = H(S^{''})$.then then they authenticate each other . Hence, membership authentication is done here.
- After authenticating each other, each member send its last calculated value in hashed form to TA. here it check for the following

$$H(C) = H(S^{'}) = H(S^{''}) \tag{6}$$

- if the equation 6 is ok. then TA authenticate members and the C is set a session key for the further communication.

V. SECURITY ANALYSIS

To illustrate and prove protocol's security and accuracy,any proposed protocol must be passed through formal and informal analysis. for formal analysis we have used most common verification tool known as AVISPA. Informal analysis is in the next subsection.

1) AVISPA: To speed up the next generation of security protocols and improve thier security, it is vital to have tools that supports the rigorous analysis of security protocols. For that it detects weaknesses and establish their correctness. In 2005, A. Armando et al. [49] came with a push button tool for the automated validation of Internet security-sensitive protocols and applications named as AVISPA.The tool is used by a protocol designer who describes a security problem in the High-Level Specification Language (HLPSL) [50]. Detailed studies of AVISPA and its execution are available in [51]. AVISPA tools provide four utilities :

- **On-the-fly Model-Checker (OFMC)**: Performs protocol falsification and bounded verification.
- **Constraint-Logic-based Attack Searcher (CL-AtSe)**: It can identify type flaws and manage message concatenation associativity.
- **SAT-based Model-Checker (SATMC)**: constructs a propositional formula encoding a bounded unrolling of the IF's specified transition relation, the initial state, and the set of conditions denoting a breach of the security properties.
- **TA4SP (Tree Automata based on Automatic Approximations for the Analysis of Security Protocols)**: Regular tree languages and rewriting are used to approximate attacker knowledge. Our protocol passed all four tests.

Proposed protocol easily passes OFMC and CL-AtSe as shown in the Fig. 7 and Fig. 8 respectively. AVISPA tools can be dowloaded at [52].

A. Informal analysis

Fig.5. explains mandatory security features for group-based authentication. Apart from supporting security features of figure .5, the proposed scheme can also support many more features such as confidentiality, scalability, and integrity.

1) **Correctness** : In authentication phase $S^{'} = \sum_{j=1}^{k} (\delta_j)$ and $S^{''} = \sum_{j=1}^{n} (F_j^1) \prod_{j=1, j\neq i}^{u} \frac{-w_i}{w_i - w_j}$ are calculated, members are said to be correct only if $S^{'} = S^{''}$. non members can not know the value of δ since it is transferred to other legitimate members in the encrypted form whose encryption key is also not known.
2) **Forward secrecy** : Because keys like $(C.T_i, F_i^1)$ can only be computed or stored by members of the secure communication group, if a member leaves the group, the departing member will be unable to access the content of future conversations.
3) **Freshness of authentication response**: In the proposed scheme, all the exchangeable value depends upon the random bivariate polynomial taken by TA. It makes $S^{'}$ and $S^{''}$ always fresh. Hence, it is impossible to impersonate

(a)Correctness: The protocol can successfully authenticate all participating users' memberships, then build a secret group key among all members, and finally, each member may successfully validate that the group key he or she computed is equal to the group key computed by other members.
(b) Freshness of authentication response: Members' authentication responses can only be utilised once during the membership authentication process. This feature can protect against a replay attack, in which an attacker replicates a recently stored authentication response in order to fail membership authentication.
(c) Freshness of group keys: Members of the key establishment generate a private group key that can only be used once for communication. This feature can prohibit attackers from using previously compromised group keys to access subsequent secure communications.
(d) Freshness of the group key authentication: Members in the group key establishment can only utilise the group key authentication messages once. This feature can protect against a replay attack, in which an attacker replays captured group key authentication communications in order to fail group key authentication.
(e) Forward secrecy of group keys: If a departing member becomes unable to access the content of any future group session's communications, forward secrecy is assured..
(f) Backward secrecy of group keys: Backward secrecy is ensured if a new member is unable to view the content of any previous session's communications.

Fig. 6. group based membership authentication process

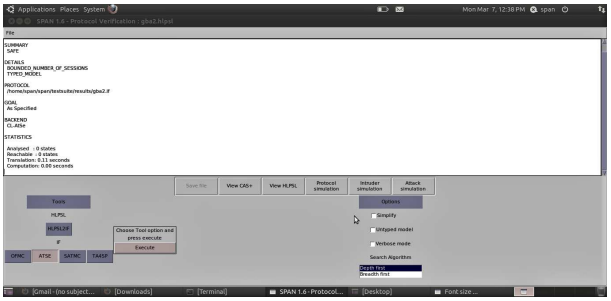


Fig. 7. CL-AtSe

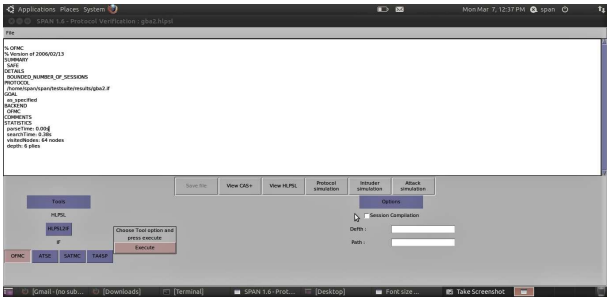


Fig. 8. OFMC

a member by recording a previously used authentication response.

4) **Backward secrecy**: If a new member joins the group, he or she will be unable to view the content of any previous conversations since the keys can only be computed by members who are participating in the secure communication.
5) **Confidentiality**: An adversary can not calculate the session keys from the values distributed on the network by trusted authority. So, the secret information always be confident.
6) **Integrity**: In our scheme, we have used hashed key to be compared that makes information protected from being deleted, modified,forged e.t.c during storage or transmission.
7) **Scalability**: IoT devices can randomly join and leave the group at any time and take part in membership and group authentication.

1) Prevention of attacks: Proposed scheme prevent following attacks:

1) **Replay attack** : The replay attack is a strategy for obtaining the secret key by capturing packets from a specific device and analyzing them. This type of behavior is prohibited under our scheme. The attack is also made even more difficult because only necessary components are used, and that too later hashed. When it concerns re-

TABLE II
CAPTION

Feature Comparison	Basic [13]	lee at.al [17]	wang et al. [48]	chein et al. [16]	Q.chang et al. [18]	Proposed scheme
Base-system	(t, n) threshold	(t, n) threshold	(t, n) threshold	(t, n) threshold,ECDP	(t, n) threshold	Bivaraite polynomial
Resist Replay attack	no	yes	yes	–	yes	yes
Resist Physical attack	no	no	yes	yes	–	yes
Resist De-synchronization	no	no	yes	no	—	yes
Resist MIM attack	no	no	yes	no	–	yes
Untracebility	no	no	yes	no	no	yes
freshness of authentication response	no	–	no	–	yes	yes
Forward secrecy	no	–	no	–	yes	yes
Backward secrecy	no	–	no	–	yes	yes

play attacks, there seem to be two approaches to consider. The first is the replay of packets in which the critical parts are swapped. Because packets should be received, processed, and then deleted by the receiver, replaying these packets serves no use other than a weak effort at a denial-of-service attack. The second approach comprises replaying data packets that are headed upstream. These packets are encrypted. Thus a replay will lead to a failure.

2) **Physical attack**:The proposed scheme is based on partitioning the secret key into pieces so that the key can be reconstructed with the fewest possible parts. The split sections are distributed among the members of the group. A secret key is broken into N pieces for a group of N members so that at least K components are required to reconstruct the secret key. Anything less will prevent the key from being reconstructed. An attacker will have great difficulty acquiring the secret key because he or she will need to know the number K and get at least K pieces. This means the attacker must capture K members out of the total of N. As a result, under our scheme, we may avoid keeping the entire secret key in the IoT device's memory.
3) **De-synchronization attack**: In this attack, an adversary blocks the communication between the devices and the trusted authority by changing the ID's of devices. Here in our proposed work, information exchanged does not depend on the device's Id, eliminating the chance of de-synchronization. attack as in [53] and [48].
4) **Man-In-The-Middle Attack** :In proposed scheme,$(C.T_i, F_i^1)$ are exchanged between each members, Real share is not accessible by having $(C.T_i, F_i^1)$. and getting secrets is much more difficult.
5) **Untraceability** : In proposed scheme, after each session, trusted authority takes different bivariate polynomial to generate different shadow secret for each member. It makes shadow secret too difficult to be traced. Furthermore, hashed form of $S^{'}$ and $S^{''}$ makes untraceability much more difficul.

B. Performance evaluation

1) Computational cost: In the authentication phase, each member needs to compute pairwise shared keys with each other members. According to Horner's rule [54], for each univariate polynomials of $t-1$ and $h-1$ degree, there is a requirement of t and h calculations. Here in the proposed protocol, $F_i^2(0)$ is constant, due to which calculation reduces to Shamir's secret sharing scheme. The computation cost is very low in comparison to the other existing schemes.

2) Communication cost: The communication cost of membership authentication is meager because all the public values associated with the specific user are broadcast. Furthermore, we have used one shared token component to encrypt others to avoid dependency on the private channel. Based on these reasons, we may conclude that it has significantly less communication cost than the existing schemes.

3) Storage cost: Each member has to store only uni-variate polynomial of modulus $\mathbb{P}$. So the storage requirement for each member is $t \log \mathbb{P}$ bits. This polynomial based modulus is far less than the public-key-based modulus [19]. Here threshold is t, to get collusion free authentication each members needs to deal with $t+1$ uni-variate polynomial at once. So overall storage cost is very less as compared to symmetric key based schemes.

4) Comparison with existing schemes: All existing schemes can support either user authentication or session key establishment. On the other hand, our protocol may enable both membership authentication and session key establishment at the same time. Furthermore, the complexity of our membership authentication is $\mathcal{O}(n)$, where n is the number of members, which is very less than the traditional one-to-one authentication scheme with complexity $\mathcal{O}(n^2)$. In Table II, we have compared with the recent scheme for grouped authentication. From there, we conclude that our scheme can deal with many attacks in the case of IoT authentication. Furthermore, in figure 6, we can see that no communication between nodes is significantly less than others.

VI. CONCLUSION

While focusing on smart cities, we need to deal with highly digitalized devices with maximum efficiency and minimum

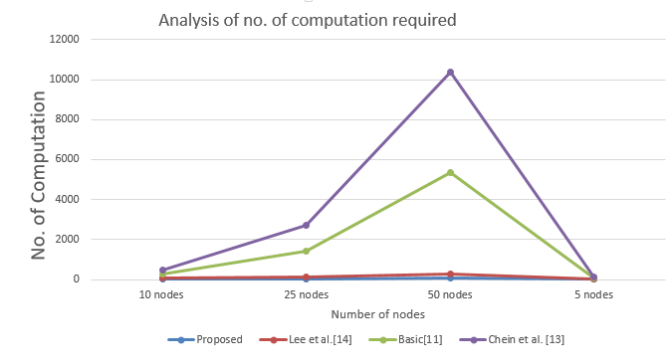


Fig. 9. Comparison of Schemes

cost in IoT infrastructure. In this article, the importance of IoT in smart cities is explained, and so its security. IoT devices perform group-based communication in smart cities, so the proposed scheme deals with group-based membership authentication. It provides recent compulsory security features and has less communication, storage, and computation costs than existing schemes.

VII. Acknowledgement

This research was carried out with the help of free, open-source arithmetic programming. We are very thankful for all open source community such as Sage Math, securitysite [55].

References

[1] S. cities council. Definitions and overviews of smart city. [Online]. Available: https://smartcitiescouncil.com/smart-cities-information-center/definitions-and-overviews

[2] V. Fernandez-Anez, "Stakeholders approach to smart cities: A survey on smart city definitions," in *International conference on smart cities*. Springer, 2016, pp. 157–167.

[3] N. C. Luong, D. T. Hoang, P. Wang, D. Niyato, D. I. Kim, and Z. Han, "Data collection and wireless communication in internet of things (iot) using economic analysis and pricing models: A survey," *IEEE Communications Surveys & Tutorials*, vol. 18, no. 4, pp. 2546–2590, 2016.

[4] A. Zanella, N. Bui, A. Castellani, L. Vangelista, and M. Zorzi, "Internet of things for smart cities," *IEEE Internet of Things journal*, vol. 1, no. 1, pp. 22–32, 2014.

[5] H. Arasteh, V. Hosseinnezhad, V. Loia, A. Tommasetti, O. Troisi, M. Shafie-khah, and P. Siano, "Iot-based smart cities: A survey," in *2016 IEEE 16th International Conference on Environment and Electrical Engineering (EEEIC)*. IEEE, 2016, pp. 1–6.

[6] N. Chai, C. Mao, M. Ren, W. Zhang, P. Poovendran, and P. Balamurugan, "Role of bic (big data, iot, and cloud) for smart cities," *Arabian Journal for Science and Engineering*, pp. 1–15, 2021.

[7] W. Iqbal, H. Abbas, M. Daneshmand, B. Rauf, and Y. A. Bangash, "An in-depth analysis of iot security requirements, challenges, and their countermeasures via software-defined security," *IEEE Internet of Things Journal*, vol. 7, no. 10, pp. 10 250–10 276, 2020.

[8] B. A. Alzahrani, "Secure and efficient cloud-based iot authenticated key agreement scheme for e-health wireless sensor networks," *Arabian Journal for Science and Engineering*, vol. 46, no. 4, pp. 3017–3032, 2021.

[9] F. Wu, X. Li, L. Xu, A. K. Sangaiah, and J. J. Rodrigues, "Authentication protocol for distributed cloud computing: An explanation of the security situations for internet-of-things-enabled devices," *IEEE Consumer Electronics Magazine*, vol. 7, no. 6, pp. 38–44, 2018.

[10] T.-Y. Wu, Z. Lee, M. S. Obaidat, S. Kumari, S. Kumar, and C.-M. Chen, "An authenticated key exchange protocol for multi-server architecture in 5g networks," *IEEE Access*, vol. 8, pp. 28 096–28 108, 2020.

[11] L. Xu and F. Wu, "A lightweight authentication scheme for multi-gateway wireless sensor networks under iot conception," *Arabian Journal for Science and Engineering*, vol. 44, no. 4, pp. 3977–3993, 2019.

[12] A. Basit, G. A. S. Sidhu, A. Mahmood, and F. Gao, "Efficient and autonomous energy management techniques for the future smart homes," *IEEE Transactions on Smart Grid*, vol. 8, no. 2, pp. 917–926, 2015.

[13] L. Harn, "Group authentication," *IEEE Transactions on computers*, vol. 62, no. 9, pp. 1893–1898, 2012.

[14] W.-T. Su, W.-M. Wong, and W.-C. Chen, "A survey of performance improvement by group-based authentication in iot," in *2016 International Conference on Applied System Innovation (ICASI)*. IEEE, 2016, pp. 1–4.

[15] A. Shamir, "How to share a secret," *Communications of the ACM*, vol. 22, no. 11, pp. 612–613, 1979.

[16] H.-Y. Chien, "Group authentication with multiple trials and multiple authentications," *Security and Communication Networks*, vol. 2017, 2017.

[17] D.-H. Lee and I.-Y. Lee, "Dynamic group authentication and key exchange scheme based on threshold secret sharing for iot smart metering environments," *Sensors*, vol. 18, no. 10, p. 3534, 2018.

[18] Q. Cheng, C. Hsu, Z. Xia, and L. Harn, "Fast multivariate-polynomial-based membership authentication and key establishment for secure group communications in wsn," *IEEE Access*, vol. 8, pp. 71 833–71 839, 2020.

[19] S. Wu, C. Hsu, Z. Xia, J. Zhang, and D. Wu, "Symmetric-bivariate-polynomial-based lightweight authenticated group key agreement for industrial internet of things," *Journal of Internet Technology*, vol. 21, no. 7, pp. 1969–1979, 2020.

[20] H. Shi, M. Fan, Y. Zhang, M. Chen, X. Liao, and W. Hu, "An effective dynamic membership authentication and key management scheme in wireless sensor networks," in *2021 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2021, pp. 1–6.

[21] L. Harn, C.-F. Hsu, Z. Xia, and J. Zhou, "How to share secret efficiently over networks," *Security and Communication Networks*, vol. 2017, 2017.

[22] L. Harn and C.-F. Hsu, "(t, n) multi-secret sharing scheme based on bivariate polynomial," *Wireless Personal Communications*, vol. 95, no. 2, pp. 1495–1504, 2017.

[23] I. Alam and A. Basit, "An extended protected secret sharing scheme," in *2019 International Conference on Electrical, Electronics and Computer Engineering (UPCON)*. IEEE, 2019, pp. 1–4.

[24] M. Shen, H. Liu, L. Zhu, K. Xu, H. Yu, X. Du, and M. Guizani, "Blockchain-assisted secure device authentication for cross-domain industrial iot," *IEEE Journal on Selected Areas in Communications*, vol. 38, no. 5, pp. 942–954, 2020.

[25] M. Wazid, A. K. Das, V. Bhat, and A. V. Vasilakos, "Lam-ciot: Lightweight authentication mechanism in cloud-based iot environment," *Journal of Network and Computer Applications*, vol. 150, p. 102496, 2020.

[26] M. Rana, A. Shafiq, I. Altaf, M. Alazab, K. Mahmood, S. A. Chaudhry, and Y. B. Zikria, "A secure and lightweight authentication scheme for next generation iot infrastructure," *Computer Communications*, vol. 165, pp. 85–96, 2021.

[27] B. D. Deebak and F. Al-Turjman, "Robust lightweight privacy-preserving and session scheme interrogation for fog computing systems," *Journal of Information Security and Applications*, vol. 58, p. 102689, 2021.

[28] U. C. Cabuk, G. Dalkilic, and O. Dagdeviren, "Comad: Context-aware mutual authentication protocol for drone networks," *IEEE Access*, 2021.

[29] C. Hsu, L. Harn, and Z. Xia, "An hss-based robust and lightweight multiple group authentication for its towards 5g," *IET Intelligent Transport Systems*, 2021.

[30] H. Xiong, J. Chen, Q. Mei, and Y. Zhao, "Conditional privacy-preserving authentication protocol with dynamic membership updating for vanets," *IEEE Transactions on Dependable and Secure Computing*, no. 01, pp. 1–1, 2020.

[31]

[32] A. Basit, N. C. Kumar, V. C. Venkaiah, S. A. Moiz, A. N. Tentu, and W. Naik, "Multi-stage multi-secret sharing scheme for hierarchical access structure," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*. IEEE, 2017, pp. 557–563.

[33] J. He and E. Dawson, "Multistage secret sharing based on one-way function," *Electronics Letters*, vol. 30, no. 19, pp. 1591–1592, 1994.

[34] ——, "Multisecret-sharing scheme based on one-way function," *Electronics Letters*, vol. 31, no. 2, pp. 93–95, 1995.

[35] B. Chor, S. Goldwasser, S. Micali, and B. Awerbuch, "Verifiable secret sharing and achieving simultaneity in the presence of faults," in *26th Annual Symposium on Foundations of Computer Science (sfcs 1985)*. IEEE, 1985, pp. 383–395.

[36] L. Harn and C.-F. Hsu, "A practical hybrid group key establishment for secure group communications," *The Computer Journal*, vol. 60, no. 11, pp. 1582–1589, 2017.

[37] ——, "A novel design of membership authentication and group key establishment protocol," *Security and Communication Networks*, vol. 2017, 2017.

[38] C.-F. Hsu, L. Harn, Y. Mu, M. Zhang, and X. Zhu, "Computation-efficient key establishment in wireless group communications," *Wireless Networks*, vol. 23, no. 1, pp. 289–297, 2017.

[39] L. Harn, C.-F. Hsu, and B. Li, "Centralized group key establishment protocol without a mutually trusted third party," *Mobile Networks and Applications*, vol. 23, no. 5, pp. 1132–1140, 2018.

[40] H. Xiong, Y. Wu, and Z. Lu, "A survey of group key agreement protocols with constant rounds," *ACM Computing Surveys (CSUR)*, vol. 52, no. 3, pp. 1–32, 2019.

[41] K. Shankar *et al.*, "Improving the security and authentication of the cloud with iot using hybrid optimization based quantum hash function," *Journal of Intelligent Systems and Internet of Things*, vol. 1, no. 2, pp. 61–1, 2021.

[42] M. Fotouhi, M. Bayat, A. K. Das, H. A. N. Far, S. M. Pournaghi, and M.-A. Doostari, "A lightweight and secure two-factor authentication scheme for wireless body area networks in health-care iot," *Computer Networks*, vol. 177, p. 107333, 2020.

[43] Z. Cui, X. Fei, S. Zhang, X. Cai, Y. Cao, W. Zhang, and J. Chen, "A hybrid blockchain-based identity authentication scheme for multi-wsn," *IEEE Transactions on Services Computing*, vol. 13, no. 2, pp. 241–251, 2020.

[44] P. Rogaway and T. Shrimpton, "Cryptographic hash-function basics: Definitions, implications, and separations for preimage resistance, second-preimage resistance, and collision resistance," in *International workshop on fast software encryption*. Springer, 2004, pp. 371–388.

[45] P. M. Mukundan, S. Manayankath, C. Srinivasan, and M. Sethumadhavan, "Hash-one: a lightweight cryptographic hash function," *IET Information Security*, vol. 10, no. 5, pp. 225–231, 2016.

[46] N. Mouha, M. S. Raunak, D. R. Kuhn, and R. Kacker, "Finding bugs in cryptographic hash function implementations," *IEEE transactions on reliability*, vol. 67, no. 3, pp. 870–884, 2018.

[47] D. Dolev and A. Yao, "On the security of public key protocols," *IEEE Transactions on information theory*, vol. 29, no. 2, pp. 198–208, 1983.

[48] A. Wang, J. Shen, L. Yan, Y. Ren, and Q. Liu, "A practical group authentication scheme for smart devices in iot," *EAI Endorsed Transactions on Internet of Things*, vol. 4, no. 15, 2018.

[49] B. D. B. Y. C. Y. C. L. C. J. . . V. L. Armando, A., "The avispa tool for the automated validation of internet security protocols and applications," *In International conference on computer aided verification*, pp. 281–285, 2005.

[50] A. B. Guide, "Hlpsl tutorial," 2006.

[51] I. Alam and M. Kumar, "A novel protocol for efficient authentication in cloud-based iot devices," *Multimedia Tools and Applications*, pp. 1–21, 2022.

[52] Avispa and span. Download avispa. [Online]. Available: http://people.irisa.fr/Thomas.Genet/span/

[53] X. Li, J. Niu, M. Z. A. Bhuiyan, F. Wu, M. Karuppiah, and S. Kumari, "A robust ecc-based provable secure authentication protocol with privacy preserving for industrial internet of things," *IEEE Transactions on Industrial Informatics*, vol. 14, no. 8, pp. 3599–3609, 2017.

[54] D. E. Knuth, "Seminumerical algorithms," *The art of computer programming*, vol. 2, 1997.

[55] secret sharing. Shamir secret shareing. [Online]. Available: https://asecuritysite.com/encryption/secret$_share$

Circuits, Systems, and Signal Processing
https://doi.org/10.1007/s00034-022-02032-4

A Novel CFDITA-Based Design of Grounded Capacitance Multiplier and Its Transpose Structure

Atul Kumar[1] · Dheeraj Singh[2] · Deva Nand[2]

Received: 6 January 2022 / Revised: 2 April 2022 / Accepted: 3 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract

A new circuit realization of capacitance multiplier based on active block, current follower differential input transconductance amplifier is introduced in the paper. The circuit employs one current follower differential input transconductance amplifier, two resistors and one capacitor. The proposed capacitance multiplier has a good range of multiplication factor (up to 30). Multiplication factor of the proposed circuit is adjustable by external resistor and by bias current as well. The comprehensive study of the proposed circuit including non-idealities and parasitic element effects has been carried out. The proposed capacitance multiplier circuit is simulated on Cadence Virtuoso using 0.18 μm gpdk CMOS technology and post layout simulation results are demonstrated to validate the working of the proposed circuit. The layout of the used active block occupies an area of 26.26 μm × 28.79 μm. The proposed circuit shows good performance against transistors mismatch, process and temperature variations. Moreover, the experimental results of the circuit using off-the-shelf integrated circuits are also carried out to demonstrate its practicality. In the last, the transpose structure of the proposed capacitance multiplier circuit is also explored. The obtained transposed structure is realized using active element namely, voltage differencing buffered amplifier.

Keywords Analog circuit · Capacitance multiplier · CFDITA · Current conveyor

✉ Dheeraj Singh
drajiitr@gmail.com

Atul Kumar
atul.nit304@gmail.com

Deva Nand
devkamboj@dce.ac.in

[1] Department of Electronics and Communication Engineering, Jaypee Institute of Information Technology, Noida, U.P. 201304, India

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

Published online: 09 May 2022

1 Introduction

In almost all electronic systems there is a need of passive components for the realization of circuit with resistor and capacitors are most common. For electronic system to work at low frequency, higher value of RC multiplication is required in order to get the higher value of time constant. A number of circuits need higher value of capacitors like sample and hold circuit, low frequency oscillators and filters [5, 7, 17]. In integrated circuits, it is recommended to use low value capacitors to avoid its complicated monolithic fabrication and larger chip area. In analog circuits, designers are interested to obtain the high value of capacitors using active blocks since the active component's size has been reduced significantly with technology advancement in comparison of passive components. A number of capacitance multiplier circuits have been implemented in the past using different active components like second-generation current controlled current conveyor (CCCII) [1], current feedback operational amplifier (CFOA) [2, 21, 31, 34, 42], operational transconductance amplifier (OTA) [3, 4, 24], second-generation current conveyor (CCII) [3, 4, 6, 13, 23], current buffered transconductance amplifier (CBTA) [8], current differencing transconductance amplifier (CDTA) [12], current operational amplifier (COA) [23], dual-X second-generation current conveyor (DXCCII) [29], current feedback transconductance amplifier (CFTA) [30], third-generation current conveyor transconductance amplifier (CCIIITA) [33], electronically controllable second-generation current conveyor (ECCII) [36], second-generation voltage conveyor (VCII) [37, 38], differential voltage current conveyor transconductance amplifier (DVCCTA) [40], fully balanced voltage differencing buffered amplifier (FB-VDBA) [41], inverting second generation current conveyor (ICCII) [43], differential voltage current conveyor (DVCC) [44] and inverting current feedback operational amplifier (ICFOA) [45]. The brief summary of recently developed capacitance multiplier circuits is given as follows. CFOA and OTA are used to realize two grounded capacitance multiplier circuits in [2]. One of the circuits employs one CFOA, one OTA along with one resistor and one capacitor. Another one is realized using one CFOA, two OTAs and one capacitor. This circuit has tunable property and also has high multiplication factor. Two capacitance multipliers using CDTA are realized in [12]. One of the circuits uses single active block CDTA, one resistor and one capacitor and another modified circuit employs one CDTA, two resistors and a single capacitor. The latter one has the advantage of lower sensitivity to temperature variations. The grounded capacitance multiplier presented in [21] utilizes two CFOAs along with three passive components including two resistors and one capacitor. This capacitance multiplier is free from the constraint of passive component matching requirement. Two CFTAs, two resistors and one capacitor are utilized for realizing the tunable grounded capacitance multiplier [30]. The circuit has the ability to eliminate parasitic capacitance. The circuit presented in [31] describes the lossy and lossless grounded capacitance multipliers utilizing CFOA. One of the capacitance multiplier circuits employs two CFOAs, two resistors and a single capacitor, whereas three CFOAs, three resistors and one capacitor are used in another circuit. In the latter one, third additional CFOA works as negative impedance converter to reduce the parasitic effects. Grounded capacitance

multiplier presented in [33] utilizes single CCIIITA along with a single capacitor only. The presented circuit is electronically tunable and consumes low power. Three circuits of capacitance multiplier are presented in [36]. Two of them employ one differential voltage buffer (DVB), one ECCII and one capacitor whereas third circuit utilizes one DVB, one ECCII, one voltage gain amplifier (VGA) and a single capacitor. The realized grounded circuit can be switched between lossy and lossless mode of operation. VCII-based grounded capacitance multiplier utilizes single VCII and a capacitor [37]. This circuit is ideal for low-voltage and low-power applications. The circuit presented in [38] uses two VCIIs, two resistors, and one capacitor for its realization. The circuit realizes a grounded capacitance multiplier which provides both positive and negative multiplication factor. Two CFOAs, one resistor and one capacitor are employed to realize a grounded capacitance multiplier in [42]. The circuit enjoys the feature of high multiplication factor. Two grounded capacitance multipliers are presented in [43]. One of the circuits employs one CCII, one voltage follower, two resistors and a single capacitor. Another circuit is based on ICCII which uses single ICCII, one inverting voltage follower, two resistors and one capacitor. The grounded capacitance multiplier presented in [45] employs one ICFOA, two resistors and a single capacitor. The circuit provides lower operating frequency range. A floating type capacitance multiplier using one CCII, one OTA, one resistor and one capacitor is introduced in [3]. The circuit is also converted to resistor-less structure by replacing the resistor by an additional OTA. Floating capacitance multiplier based on CCII and OTA in [4] employs one CCII, four OTAs and one capacitor. The circuit has wide tuning range and also consumes low power. Two floating capacitance multipliers using CFOA are reported in [34]. One of them uses two CFOAs, two resistors and a capacitor. Another circuit uses two CFOAs, two OTAs and one capacitor. Both circuits provide negative multiplication factor. The capacitance multiplier circuit of [41] is realized using single active element FB-VDBA and two passive components including one resistor and one capacitor. The circuit realizes a floating type capacitance multiplier and the capacitance value is electronically adjustable via bias current.

The purpose of this work is to demonstrate a grounded capacitance multiplier circuit based on a single current follower differential input transconductance amplifier (CFDITA). The proposed circuit employs one CFDITA, two resistors and one capacitor. The proposed circuit is simpler as it uses single CFDITA which has only 13 MOS transistors in its internal structure. The proposed circuit's transposition structure is also investigated. The proposed circuit is verified using post layout simulations. The experimental verification of the proposed circuit is also done.

This paper is organized in nine sections. Section 2 describes the active block, CFDITA. Section 3 discusses the proposed circuit of capacitance multiplier whereas Sect. 4 comprises its parasitic and non-ideal analysis. Section 5 gives post layout simulation results. Experimental verification of the circuit using commercially available ICs is given in Sect. 6. Section 7 introduces transpose structure of the proposed capacitance multiplier. Section 8 summarizes the comparison with the relevant works. Section 9 concludes the work.

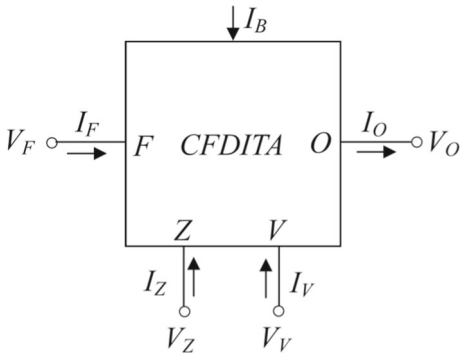


Fig. 1 Symbol of the CFDITA

2 Description of CFDITA

Current-mode active elements offer several advantages over voltage-mode counterparts, including larger bandwidth, higher slew rate, wider dynamic range and low power consumption [32]. Current conveyor and its variants are the widely used current-mode active elements. Electronically tunable current conveyors such as CDTA [10], VDBA [11], dual-X current conveyor transconductance amplifier (DXCCTA) [14] and current differencing cascaded transconductance amplifiers (CDCTA) [18] offer the additional feature of electronic tunability of transconductance. CFDITA is also one of the electronically tunable current-mode active elements, which has been used in a variety of applications including an instrumentation amplifier, a square and triangular waveform generator and a bistable multivibrator demonstrating its versatility [15, 16, 25, 26]. Figure 1 illustrates CFDITA's symbol. Current follower stage, in which current at Z terminal follows the current at F terminal, is followed by a differential input transconductance amplifier which produces output current in proportion to the difference between voltages of Z and V terminals. The ideal relationships between CFDITA's terminal currents and voltages are represented in (1) [25].

$$\begin{bmatrix} I_V \\ I_Z \\ I_O \\ V_F \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -g_m & g_m & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} V_V \\ V_Z \\ V_O \\ I_F \end{bmatrix} \tag{1}$$

where g_m denotes CFDITA's transconductance.

3 Proposed Circuit of Capacitance Multiplier Using CFDITA

The proposed capacitance multiplier circuit, which has a simpler circuit configuration, is shown in Fig. 2. The circuit structure consists of a single CFDITA, one grounded and one floating resistor and one capacitor. The circuit is analyzed using CFDITA's ideal terminal relationships, which are given in (1). The current entering in terminal F is given as follows.

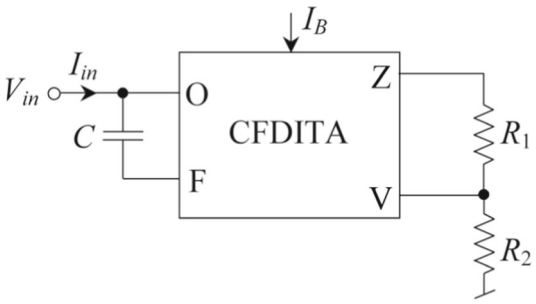


Fig. 2 The proposed capacitance multiplier circuit

$$I_F = V_{in} sC \tag{2}$$

The input current is given as follows.

$$I_{\text{in}} = I_{\text{F}} - I_{\text{O}} = I_{\text{F}} - g_{\text{m}}(V_{\text{Z}} - V_{\text{V}}) \tag{3}$$

The voltages developed at Z and V terminals are given as follows.

$$V_{\text{Z}} = -(R_1 + R_2) I_{\text{F}} \tag{4}$$

$$V_{\text{V}} = -R_2 I_{\text{F}} \tag{5}$$

After putting the values of I_{F}, V_{Z} and V_{V} in (3), the expression of input impedance is obtained as given below.

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{1}{sC(1 + g_m R_1)} \tag{6}$$

From (6), the equivalent capacitance can be written in terms of input capacitance as follows.

$$C_{eq} = C(1 + g_m R_1) \tag{7}$$

The multiplication factor (MF) is given as follows.

$$MF = 1 + g_m R_1 \tag{8}$$

MF is dependent on transconductance and external resistor. Therefore, MF can be tuned by adjusting the external resistor value and can also be tuned by adjusting the bias current.

4 Non-idealities and Parasitic Element Effects

4.1 Non-ideal Study

CFDITA's non-ideal terminal relationships can be described as given below.

$$I_V = 0, I_Z = \alpha I_F, I_O = g_m \gamma (V_Z - V_V) \tag{9}$$

where α represents the non-ideal current transfer gain from terminal F to terminal Z and deviation in transconductance value is denoted by γ. Non-idealities represented by (9) are utilized to reanalyze the circuit. The expressions in (4) and (5) are now modified as follows.

$$V_{\mathrm{Z}} = -\alpha (R_1 + R_2) I_{\mathrm{F}} \tag{10}$$

$$V_{\mathrm{V}} = -\alpha R_2 I_{\mathrm{F}} \tag{11}$$

The expression of modified input impedance is given below.

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{1}{sC(1 + \alpha \gamma g_m R_1)} \tag{12}$$

The equivalent capacitance of the proposed capacitance multiplier in terms of input capacitance is represented as follows.

$$C_{eq} = C(1 + \alpha \gamma g_m R_1) \tag{13}$$

The modified MF is given as follows.

$$MF = 1 + \alpha \gamma g_m R_1 \tag{14}$$

The values of non-idealities α and γ differ from unity only at higher frequencies. Therefore MF in (14) is affected by non-idealities at higher frequencies only.

4.2 Parasitic Effects

Parasitic element effects can be studied by considering all the parasitic associated with all the terminals of CFDITA. At terminal F, a small parasitic resistance R_{F} is present and at terminals Z, V and O the parallel combinations of R_i and C_i (where $i = Z$, V and O, respectively) are present. Figure 3 depicts the suggested capacitance multiplier circuit with parasitic elements.The following expression of input impedance is obtained from a regular examination of the circuit shown in Fig. 3.

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{R_F + \frac{1}{sC}}{1 + g_m K} \tag{15}$$

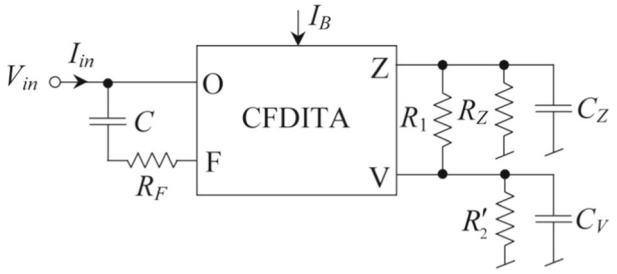


Fig. 3 Capacitance multiplier circuit with parasitic elements

where $K = \frac{R_1}{1+\frac{R_1}{R_Z}(sR_ZC_Z+1)+\frac{R_2'(sR_ZC_Z+1)}{R_Z(sR_2'C_V+1)}}$; $R_2' = R_2 \| R_V$.

For the proper operation of the proposed capacitance multiplier circuit, the following conditions are to be followed.

(i) $R_F \ll \frac{1}{sC}$ which leads to the following limitation on operating frequency:$\omega_1 \ll \frac{1}{R_FC}$

(ii) $sR_ZC_Z \ll 1$ which leads to the following limitation on operating frequency: $\omega_2 \ll \frac{1}{R_1'C_z}$

(iii) $sR_2'C_V \ll 1$ which leads to the following limitation on operating frequency:$\omega_3 \ll \frac{1}{R_2'C_V}$

The limitation of operating frequency of the proposed capacitance multiplier circuit is given as follows.

$$\omega < \min(\omega_1, \omega_2, \omega_3) \tag{16}$$

5 Post Layout Simulation Results

The functionality of the proposed capacitance multiplier has been verified through post layout simulations which are carried on Cadence Virtuoso using 0.18 μm gpdk technology. The CMOS of CFDITA is depicted in Fig. 4. Transistor dimensions in micrometer (μm) are as follows: $(W/L)_{1,2}$ – 0.14/0.18, $(W/L)_{3,4}$ – 1.44/0.18, $(W/L)_{5,6}$ – 3.6/0.18, $(W/L)_{7,8,9}$ – 16.2/0.18, $(W/L)_{10,11}$ – 9/0.18 and $(W/L)_{12,13}$ – 1/0.18. In simulation results, the bias current I_B is kept at 180 μA, bias voltage V_{BB} is − 0.7 V and power supply voltages are set to ± 1.25 V. The transconductance g_m obtained for the provided bias current is 1 mS. The passive components $C = 100$ pf, $R_1 = 10$ kΩ and $R_2 = 1$ kΩ are chosen to yield a MF of 11. Figure 5 shows the circuit's AC response for input impedance and shows good linearity in frequency range from 10 kHz to 3 MHz. MF calculated from the AC response of Fig. 5 is found to be 10.82 (1.63% error) which is closer to the theoretical value. The proposed capacitance multiplier's electronic tunability is demonstrated in Fig. 6 which shows the AC responses of input

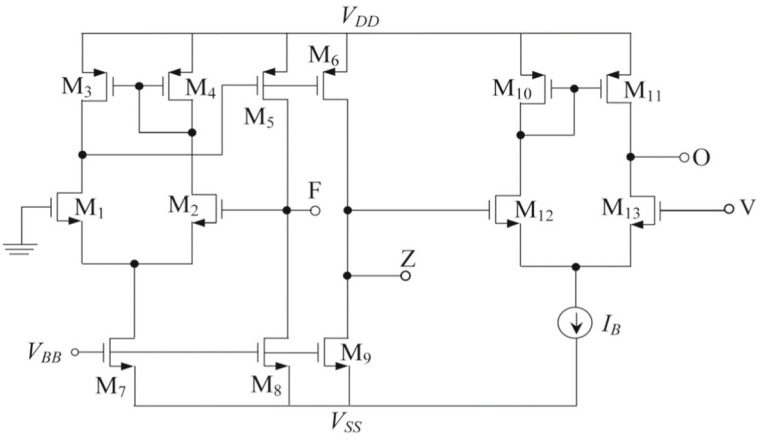


Fig. 4 CFDITA's CMOS realization [15]

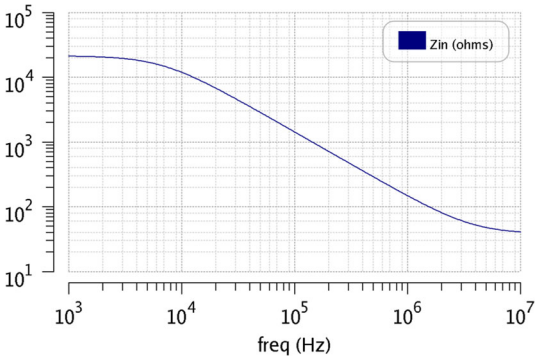


Fig. 5 AC response for input impedance

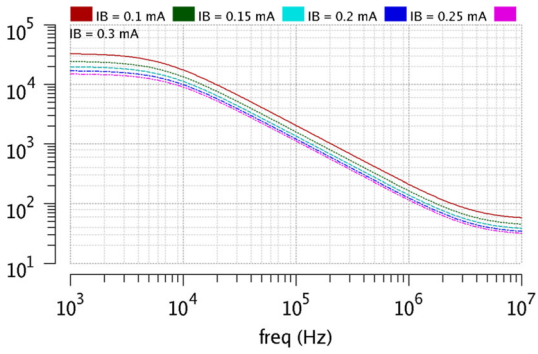


Fig. 6 Input impedance for different values of I_B

Table 1 Adjustment of MF using bias current where $R_1 = 10$ kΩ, $R_2 = 1$ kΩ,$C_{in} = 100$ pF

I_B (μA)	g_m (mS)	C_{eq} (pF)	Multiplication factor
100	0.7	750	7.5
150	0.9	964	9.64
200	1.1	1136	11.36
250	1.2	1270	12.7
300	1.3	1383	13.83

impedance for different values of bias current (variation of I_B is 100 μA to 300 μA at 50 μA step size). In addition, Table 1 shows the values of MF for various bias current levels. Moreover, the tunability of capacitance through external resistor R_1 is examined. The variations of ideal and simulated values of MF when resistor R_1 is varied from 5 kΩ to 25 kΩ are plotted in Fig. 7. The MF values for different values of resistor R_1 in Fig. 7 are also listed in Table 2. The maximum value of MF of the proposed capacitance multiplier within 10% error is found 30. Figure 8 shows the input voltage and obtained current waveforms at a frequency of 1 MHz with maximum magnitude of input voltage that can be applied to the circuit is 100 mV.

To verify the robustness against temperature, AC response is obtained by sweeping the temperature from − 50 to 50 °C as shown in Fig. 9. The proposed circuit, as shown in Fig. 9, is less impacted by temperature variations.

The circuit performance against process and mismatch variations for all transistor is examined through Monte Carlo analysis for 100 runs where the sampling method used

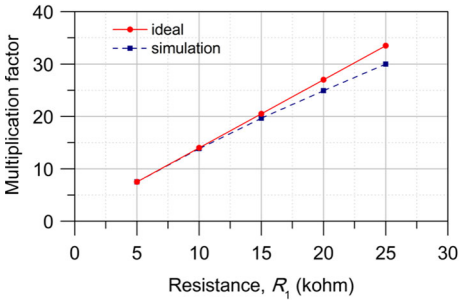


Fig. 7 MF tuning via resistor, R_1

Table 2 Adjustment of MF via R_1 where $R_2 = 1$ kΩ, $I_B = 300$ μA, $C_{in} = 100$ pF

R_1 (kΩ)	C_{eq} (pF)	Multiplication factor
5	750	7.5
10	1383	13.83
15	1964	19.64
20	2490	24.9
25	3000	30

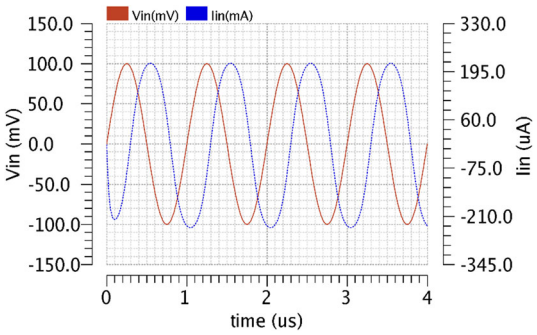


Fig. 8 Input Voltage and current waveform for sinusoidal input

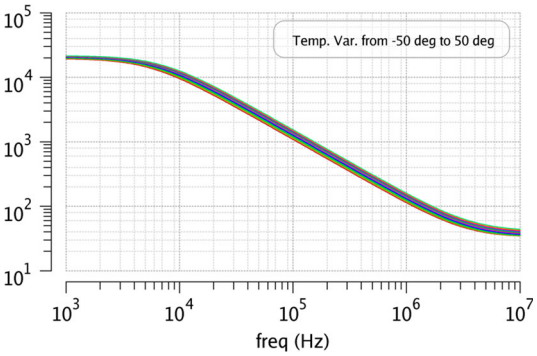


Fig. 9 Input impedance for temperature variation

is random. Monte Carlo simulations of AC responses for input impedance is shown in Fig. 10. Additionally, the corner analysis of the proposed circuit for typical, fast–fast (FF), fast-slow (FS), slow-fast (SF) and slow-slow (SS) transistor models are carried out as shown in Fig. 11. It is observed from Fig. 11 that the proposed capacitance multiplier is less affected by SF and SS models as compare to FF and FS models.

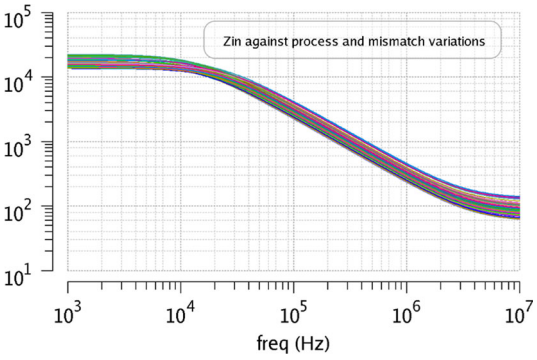


Fig. 10 Monte Carlo simulation results showing AC responses of input impedance

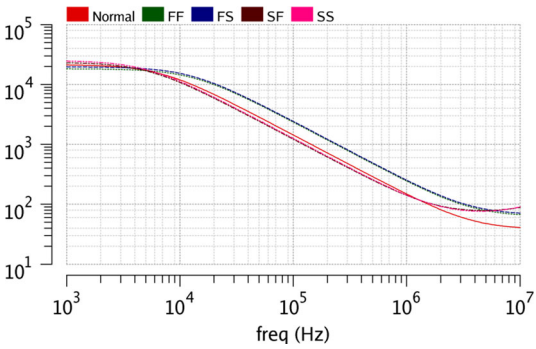


Fig. 11 AC responses of input impedance for different corners

6 Experimental Verification

The proposed capacitance multiplier is also verified experimentally. The practical realization of CFDITA using off-the-shelf ICs AD844 and LM13700 is depicted in Fig. 12a. The input current of the proposed capacitance multiplier is measured in terms of voltage as per the scheme shown in Fig. 12b where CM stands for capacitance multiplier. Figure 13 depicts the proposed capacitance multiplier's experimental setup. The power supplies of $\pm$ 10 V, passive component values $C = 100$ pf, $R_1 = 5$ kΩ, $R_2 = 1$ kΩ and bias current $I_B = 70$ μA ($g_m = 1$ mS) are used in the experimental setup. The applied input voltage is 2 V in magnitude. The waveforms of the applied input voltage and observed input current at 10 kHz frequency are shown in Fig. 14a and Lissajous pattern between applied input voltage and observed input current is shown in Fig. 14b. Lissajous pattern confirms a 90° phase difference between applied input voltage and observed input current. In Fig. 14, MF value is obtained as 6.12 which is closer to the theoretical one. Additionally, waveforms of applied input voltage and observed input current at 5 kHz and 20 kHz are shown in Fig. 15.

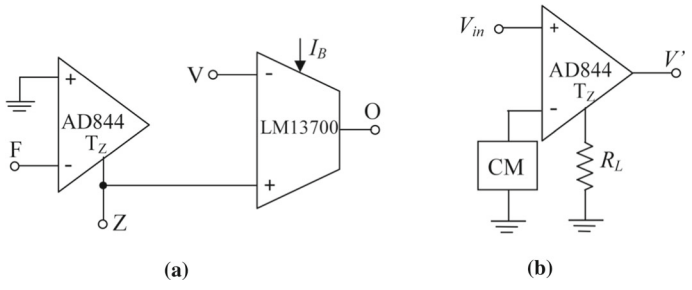


Fig. 12 (a) Practical implementation of CFDITA (b) Input current measurement in terms of voltage

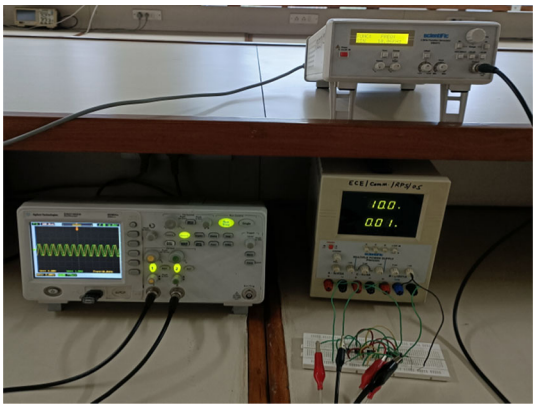
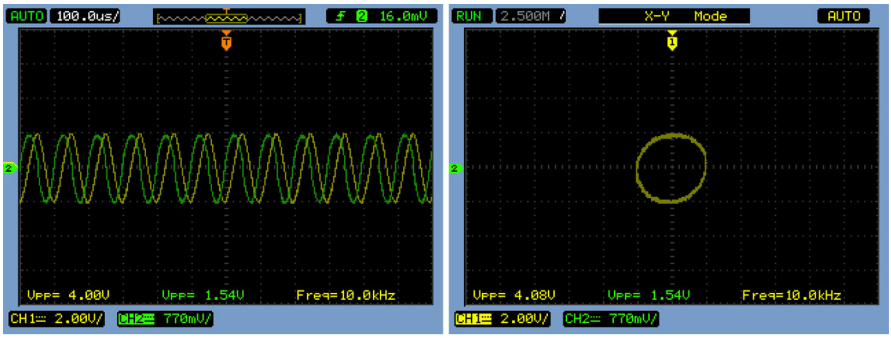
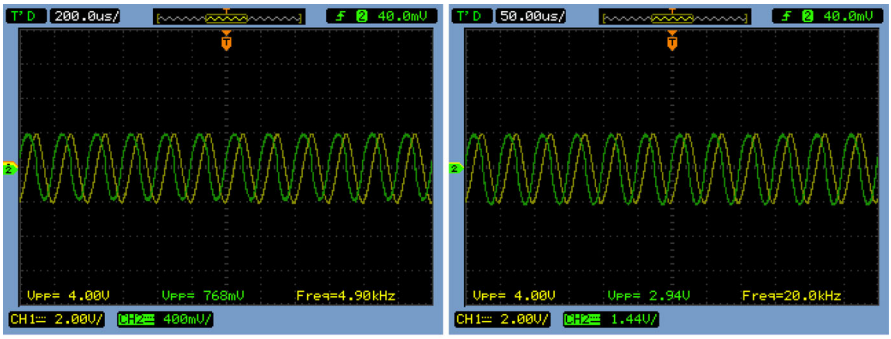


Fig. 13 Experimental setup of the proposed capacitance multiplier



(a) **(b)**

Fig. 14 (a) Waveforms of input voltage and input current at 10 kHz (b) Lissajous pattern



(a) **(b)**

Fig. 15 Waveforms of input voltage and input current at (a) 5 kHz (b) 20 kHz

7 The Transpose Structure of the Proposed CFDITA-Based Capacitance Multiplier

Transposition is a method by which a current-mode circuit structure can be derived from a voltage-mode circuit structure and vice versa. Initially, transposition concept was given in [9] in which the application of transposition to obtain a current transfer function from voltage transfer function is demonstrated. The idea to determine the transpose of any multi-terminal active element was first given in [9]. In the literature, numerous works on transposition concept have now been reported [27, 39].

In this section, transpose structure of CFDITA is derived and shown in Fig. 16. The procedure to obtain transpose structure is given as follows. The CFDITA's terminals relationship matrix can be represented as follows.

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \bar{V}_4 \end{bmatrix} = \left[\begin{array}{ccc|c} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -g_m & g_m & 0 & 0 \\ \hline \bar{0} & \bar{0} & 0 & 0 \end{array}\right] \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \bar{I}_4 \end{bmatrix} \tag{17}$$

Another way to represent matrix (17) is given below.

$$\begin{bmatrix} I_M \\ V_N \end{bmatrix} = \begin{bmatrix} Y_{11} & Y_{12} \\ Y_{21} & Y_{22} \end{bmatrix} \begin{bmatrix} V_M \\ I_N \end{bmatrix} \tag{18}$$

where $I_M = \begin{bmatrix} I_1 & I_2 & I_3 \end{bmatrix}^T$, $I_N = [I_4]^T$, $V_M = \begin{bmatrix} V_1 & V_2 & V_3 \end{bmatrix}^T$ and $V_N = [V_4]^T$

The transpose of the matrix in (18) may be given as follows [39].

$$\begin{bmatrix} I_M \\ V_N \end{bmatrix} = \begin{bmatrix} Y_{11}^T & -Y_{21}^T \\ -Y_{12}^T & Y_{22}^T \end{bmatrix} \begin{bmatrix} V_M \\ I_N \end{bmatrix} \tag{19}$$

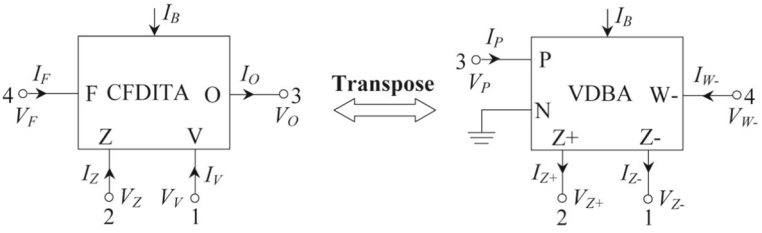


Fig. 16 CFDITA and its transpose (VDBA)

Using (19) the resultant matrix of the transposed active element can be represented as follows.

$$\begin{bmatrix} I_1 \\ I_2 \\ I_3 \\ \bar{V}_4 \end{bmatrix} = \left[\begin{array}{ccc|c} 0 & 0 & -g_m & 0 \\ 0 & 0 & g_m & 0 \\ \underline{0} & \underline{0} & \underline{0} & \underline{0} \\ 0 & -1 & 0 & 0 \end{array}\right] \begin{bmatrix} V_1 \\ V_2 \\ V_3 \\ \bar{I}_4 \end{bmatrix} \tag{20}$$

As per the above matrix equation, the relationship between currents and voltages are given as follows

$$I_1 = -g_m V_3,\ I_2 = g_m V_3,\ V_4 = -V_2 \tag{21}$$

The relationship $I_1 = -g_m V_3$ and $I_2 = g_m V_3$ represents the balanced output operational transconductance amplifier and $V_4 = -V_2$ defines the inverting voltage buffer. Voltage differencing buffered amplifier (VDBA) is an active element which is the combination of balanced output operational transconductance amplifier and inverting voltage buffer. So the transpose of CFDITA is VDBA as depicted in Fig. 16. The VDBA's terminal characteristics are listed below.

$$I_{Z-} = -g_m V_P,\ I_{Z+} = g_m V_P,\ V_{W-} = -V_{Z+} \tag{22}$$

The transpose structure of the proposed capacitance multiplier is now obtained by replacing the CFDITA in Fig. 2 by its transpose i.e., VDBA. The transpose structure of the proposed capacitance multiplier is shown in Fig. 17. The routine analysis of circuit provides the following expression of input impedance.

$$Z_{in} = \frac{V_{in}}{I_{in}} = \frac{1}{sC(1 + g_m R_1)} \tag{23}$$

It is observed from (23) that the expression of input impedance obtained from the transpose structure is same as obtained in (6) from the proposed circuit of Fig. 2. The circuit of Fig. 17 is further simulated to check its functionality. The CMOS structure of VDBA given in [35] is used to simulate the circuit. The passive components used are $C = 100$ pf, $R_1 = 10$ kΩ, $R_2 = 1$ kΩ. The AC responses for the input impedance

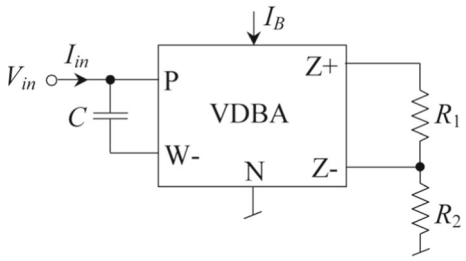


Fig. 17 Capacitance Multiplier using VDBA

at different bias current values are shown in Fig. 18. The linearity of the proposed capacitance multiplier observed from Fig. 18 lies in the frequency range of 10 Hz to 10 MHz. Additionally, the values of MF for different values of bias current are listed in Table 3. The tuning of MF with the help of resistor is shown in Fig. 19. The values of MF for different values of resistor are also listed in Table 4. The maximum value of MF of the proposed capacitance multiplier of Fig. 17 within 10% error range is found 28.32. The thermal stability is further examined by varying the temperature from − 50 to 50 ℃ as shown in Fig. 20 which shows that the circuit has very good stability with respect to temperature variation. Moreover, the performance of VDBA-based capacitance multiplier circuit against process and mismatch variations is examined. Figure 21 shows the Monte Carlo simulations for 100 runs (with random sampling method) for AC response of input impedance. It is clearly seen from Fig. 21 that

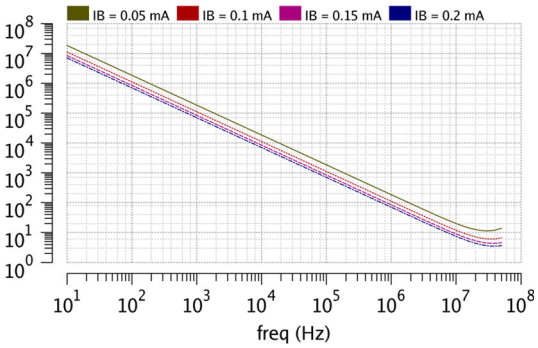


Fig. 18 Input impedance tuning via I_B

Table 3 Adjustment of MF using bias current where $R_1 = 10\,\text{k}\Omega$, $R_2 = 1\,\text{k}\Omega$, $C_{in} = 100\,\text{pF}$

I_B (μA)	g_m (mS)	C_{eq} (pF)	Multiplication factor
50	0.78	864	8.64
100	1.36	1433	14.33
150	1.8	1894	18.94
200	2.2	2273	22.73

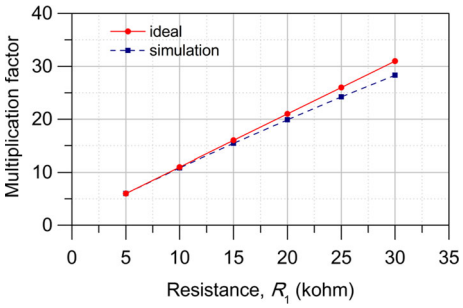


Fig. 19 The tuning of MF via resistor R_1

Table 4 Adjustment of MF via R_1 where $R_2 = 1\ \text{k}\Omega$, $I_B = 68\ \mu\text{A}$, $C_{in} = 100\ \text{pF}$

R_1 (kΩ)	C_{eq} (pF)	Multiplication factor
5	600	6
10	1082	10.82
15	1545	15.45
20	1990	19.9
25	2420	24.2
30	2832	28.32

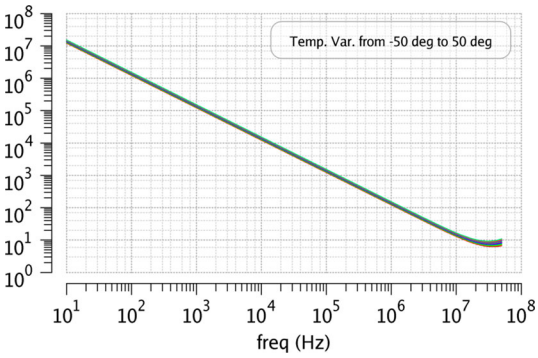


Fig. 20 AC response of input impedance for different temperature values

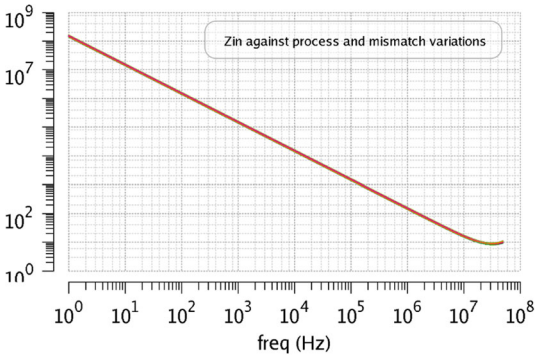


Fig. 21 Monte Carlo simulation results showing AC responses of input impedance

the proposed VDBA-based circuit is affected negligibly by process and mismatch variations.

8 Comparison of the Performance Parameters and Discussion

Table 5 summarizes the comparison of the proposed capacitance multiplier circuit with the other relevant grounded type capacitance multiplier circuits. From the comparison table, it is evident that the proposed circuit realization offers the following

Table 5 Performance comparison of proposed capacitance multiplier with existing grounded capacitance multiplier circuits

References	Type of active element	No. of active elements	Count of R, C	Count of transistors	Electronic tuning	Technology	Power supply	Post layout simulation	Experimental verification	Transpose obtained
[1]	CCCII +	3	1,1	NA	Yes	NA	± 2.5 V	No	No	No
[2]	CFOA, OTA	1,1	1,1	–	Yes	–	± 5 V	No	Yes	No
[2]	CFOA, OTA	1,2	0,1	–	Yes	–	± 5 V	No	Yes	No
[6]	CCII + , CCII-	1,1	2,2	32	No	0.13 µm	± 1 V	No	Yes	No
[12]	CDTA	1	1,1	–	Yes	–	–	No	Yes	No
[12]	CDTA	1	2,1	–	No	–	–	No	Yes	No
[13]	CCII-	1	0,1	12	Yes	1.2 µm	5 V	No	No	No
[13]	CCII + , CCII-	1,1	2,1	22	No	1.2 µm	5 V	No	No	No
[21]	CFOA	2	2,1	–	No	–	± 9 V	No	Yes	No
[23]	CCII-, COA	1,1	2,1	10	No	0.5 µm	1.5 V	No	No	No
[29]	DXCCII	1	0,1	20–22	Yes	0.35 µm	± 1.65 V	No	No	No
[30]	CFTA	2	2,1	54	Yes	0.13 µm	± 0.75 V	No	Yes	No
[31]	CFOA	2	2,1	56	No	0.13 µm	± 0.75 V	No	No	No
[31]	CFOA	3	3,1	84	No	0.13 µm	± 0.75 V	No	No	No
[33]	CCIIITA	1	0,1	74	Yes	0.18 µm	± 1.5 V	No	Yes	No
[36]	DVB, ECCII-	1,1	0,1	–	Yes	–	–	No	Yes	No
[36]	DVB, ECCII-, VGA	1,1,1	0,1	–	Yes	–	–	No	Yes	No
[37]	VCII	1	0,1	12	Yes	0.18 µm	± 0.3 V	No	No	No

Table 5 (continued)

References	Type of active element	No. of active elements	Count of R, C	Count of transistors	Electronic tuning	Technology	Power supply	Post layout simulation	Experimental verification	Transpose obtained
[38]	VCII	2	2,1	32	No	0.35 μm	± 1.65 V	No	Yes	No
[42]	CFOA	2	1,1	–	No	–	–	No	No	No
[43]	CCII-, VF	1,1	2,1	15	No	0.18 μm	± 0.9 V	No	No	No
[43]	ICCII, IVF	1,1	2,1	22	No	0.18 μm	± 0.9 V	No	No	No
[45]	ICFOA	1	2,1	46	No	0.18 μm	± 1.25 V	No	No	No
Proposed work	CFDITA	1	2,1	13	Yes	0.18 μm	± 1.25 V	Yes	Yes	Yes

NA not applicable, – not given

beneficial features: simpler circuit structure as it uses single active element, less count of transistors (only 13), electronic tunability of capacitance, verified using post layout simulation results and experimental results as well. Additionally transpose structure of the CFDITA-based capacitance multiplier circuit has also been explored. The obtained transpose structure of capacitance multiplier is based on VDBA. The VDBA-based capacitance multiplier circuit is also simpler, uses single active element and has electronically tunable feature. Moreover, it is less sensitive to the temperature and process variation effects.

Evolutionary algorithms are the well-known global optimizers to deal with multi-objective problems. Numerous examples of evolutionary algorithms, such as optimization of OTA [19] and CCII [22], linearization of transconductance of OTA [20] and improvement in quality factor of capacitance multiplier [28] are available in the literature. The proposed circuit can also be optimized by using evolutionary algorithms to provide the improved performance parameters at low power consumption. However, optimization of the proposed circuit is not included here due to brevity reasons.

9 Conclusion

A novel capacitance multiplier circuit using single CFDITA, two resistors and one capacitor has been realized in this work. The transpose of the proposed capacitance multiplier is also explored. The transpose structure uses one VDBA, two resistors and one capacitor. Both structures provide the same expression of driving point input impedance. The capacitance of the proposed circuits can be electronically tuned by adjusting bias current. Additionally, it is also tunable by external resistor. To confirm the functionality of the proposed structures, the simulation results are presented. Moreover, CFDITA based capacitance multiplier has also been verified experimentally using off-the-shelf ICs, AD844 and LM13700.

Data availablity Data sharing not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. M.T. Abuelma'atti, N.A. Tasadduq, Electronically tunable capacitance multiplier and frequency dependent negative-resistance simulator using the current controlled current conveyor. Microelectron. J. **30**(9), 869–873 (1999)
2. M.A. Al-Absi, M.T. Abuelmaatti, A novel tunable grounded positive and negative impedance multiplier. IEEE Trans. Circuits Syst. II **66**(6), 924–927 (2019)
3. M.A. Al-Absi, M.T. Abulemaatti, A tunable floating impedance multiplier. Arabian J. Sci. Eng. **44**, 7085–7089 (2019)
4. M.A. Al-Absi, A.A. Al-Khulaifi, A new floating and tunable capacitance multiplier with large multiplication factor. IEEE Access. **7**, 120076–120081 (2019)
5. M.A. Al-Absi, E.S. Al-Suhaibani, M.T. Abuelma'atti, A new compact CMOS C-multiplier. Analog Integr. Circuits Signal Process. **90**(3), 653–658 (2017)
6. H. Alpaslan, E. Yuce, Bandwidth expansion methods of inductance simulator circuits and voltage-mode biquads. J. Circuits Syst. **20**(3), 557–572 (2011)

7. R. Aparicio, A. Hajimiri, Capacity limits and matching properties of integrated capacitors. IEEE J. Solid State Circuits **37**(3), 384–393 (2002)
8. U.E. Ayten, M. Sagbas, N. Herencsar, J. Koton, Novel floating general element simulators using CBTA. Radioengineering. **21**(1), 11–19 (2012)
9. B.B. Bhattacharyya, M.N.S. Swamy, Network transposition and its application in synthesis. IEEE Trans. Circuit Theory **18**(3), 394–397 (1971)
10. D. Biolek, CDTA-building block for current-mode analog signal processing. In Proceedings of the ECCTD'03, Krakow, Poland, vol. 3, pp. 397–400 (2003)
11. D. Biolek, R. Senani, V. Biolkova, Z. Kolka, Active elements for analog signal processing: classification, review, and new proposals. Radioengineering. **17**(4), 15–32 (2008)
12. D. Biolek, J. Vavra, A. Keskin, CDTA-based capacitance multipliers. Circuits Syst. Signal Process. **38**(4), 1466–1481 (2019)
13. G.D. Cataldo, G. Ferri, S. Pennisi, Active capacitance multipliers using current conveyors. In IEEE international symposium on circuits and systems (ISCAS), pp. 343–346 (1998)
14. B. Chaturvedi, A. Kumar, DXCCTA: a new active element. In IEEE 1st international conference on power electronics, intelligent control and energy systems (ICPEICES), pp. 1–6 (2016)
15. B. Chaturvedi, A. Kumar, CMOS MO-CFDITA based fully electronically controlled square/triangular wave generator with adjustable duty cycle. IET Circuits Devices Syst. **12**(6), 817–826 (2018)
16. B. Chaturvedi, A. Kumar, Electronically tunable current-mode instrumentation amplifier with high CMRR and wide bandwidth. AEU-Int. J. Electron. Commun. **92**, 116–123 (2018)
17. J. Choi, J. Park, W. Kim, K. Lim, J. Laskar, High multiplication factor capacitor multiplier for an on-chip PLL loop filter. Electron. Lett. **45**(5), 239–240 (2009)
18. E.T. Cuautle, L.G. De la Fraga, K. Phanrattanachaic, K. Pitaksuttayaprot, CDCTA and OTA realization of a multi-phase sinusoidal oscillator. IETE Tech. Rev. **32**(6), 497–504 (2015)
19. E.T. Cuautle, A.C. Sanabria-Borbon, Optimising operational amplifiers by evolutionary algorithms and gm/Id method. Int. J. Electron. **103**(10), 1665–1684 (2016)
20. L.G. De la Fraga, E.T. Cuautle, Linearizing the transconductance of an OTA through the optimal sizing by applying NSGA-II. In 15th International conference on synthesis, modeling, analysis and simulation methods and applications to circuit design (SMACD), IEEE, pp. 1–9 (2018)
21. M. Dogan, E. Yuce, A new CFOA based grounded capacitance multiplier. AEU-Int. J. Electron. Commun. **115**, 153034 (2020)
22. M.E. Fakhfakh, T. Cuautle, D. Moro, M. Loulou, Optimized CCII–based tuneable filter and oscillator using minimum number of passive elements. In: 5th International conference: sciences of electronic, technologies of information and telecommunications (SETIT), IEEE (2009)
23. G. Ferri, S. Pennisi, A 1.5-V current-mode capacitance multiplier. In Proceedings of the tenth international conference on microelectronics, pp. 9–12 (1998)
24. I.A. Khan, M.T. Ahmed, OTA-based integrable voltage/current-controlled ideal C-multiplier. Electron. Lett. **22**(7), 365–366 (1986)
25. A. Kumar, B. Chaturvedi, Novel CMOS CFDITA and its application as electronically tunable bistable multivibrator. In: International conference on signal processing and communication (ICSC), pp. 374–379 (2016)
26. A. Kumar, B. Chaturvedi, Fully electronically controllable Schmitt trigger circuit with dual hysteresis. Electron. Lett. **53**(7), 459–461 (2017)
27. A. Kumar, B. Chaturvedi, J. Mohan, Possible active elements transposition and applications. In International conference on signal processing and communications (ICSC), pp. 271–274 (2019)
28. J. López-Arredondo, E.T. Cuautle, L.G. De la Fraga, High-Q and wide-bandwidth capacitor multiplier optimized by NSGA-II. IETE J. Res. **65**(5), 661–666 (2019)
29. I. Myderrizi, A. Zeki, Electronically tunable DXCCII-based grounded capacitance multiplier. AEU-Int. J. Electron. Commun. **68**(9), 899–906 (2014)
30. E. Özer, Electronically tunable CFTA based positive and negative grounded capacitance multipliers. AEU-Int. J. Electron. Commun. **134**, 153685 (2021)
31. E. Ozer, M.E. Basak, F. Kacar, Realizations of lossy and lossless capacitance multiplier using CFOAs. AEU-Int. J. Electron. Commun. **127**, 153444 (2020)
32. A.S. Sedra, K. Smith, A second-generation current conveyor and its applications. IEEE Trans. Circuit Theory **17**(1), 132–134 (1970)
33. S. Singh, Jatin, N. Pandey, R. Pandey, Electronically tunable grounded capacitance multiplier. IETE J. Res. (2020). https://doi.org/10.1080/03772063.2020.1739573

34. S. Singh, N. Pandey, R. Pandey, CFOA based negative floating capacitance multiplier, in *Advances in Electronics Engineering*. ed. by G. Chandrarathne (Springer, Singapore, 2020), pp. 231–244
35. R. Sotner, J. Jerabek, N. Herencsar, Voltage differencing buffered/inverted amplifiers and their applications for signal generation. Radioengineering. **22**(2), 490–504 (2013)
36. R. Sotner, J. Jerabek, L. Polak, J. Petrzela, Capacitance multiplier using small values of multiplication factors for adjustability extension and parasitic resistance cancellation technique. IEEE Access. **8**, 144382–144392 (2020)
37. V. Stornelli, L. Safari, G. Barile, G. Ferri, A new extremely low power temperature insensitive electronically tunable VCII-based grounded capacitance multiplier. IEEE Trans. Circuits Syst. II Express Briefs **68**(1), 72–76 (2021)
38. V. Stornelli, L. Safari, G. Barile, G. Ferri, A new VCII based grounded positive/negative capacitance multiplier. AEU-Int. J. Electron. Commun. **137**, 153793 (2021)
39. M.N.S. Swamy, Transpose of a multiterminal element and applications. IEEE Trans. Circuits Syst. II **57**(9), 696–700 (2010)
40. W. Tangsrirat, Floating simulator with a single DVCCTA. Indian J. Eng. Mater. Sci. **20**(2), 79–86 (2013)
41. W. Tangsrirat, O. Channumsin, Tunable floating capacitance multiplier using single fully balanced voltage differencing buffered amplifier. J. Commun. Technol. Electron. **64**(8), 797–803 (2019)
42. R. Verma, N. Pandey, R. Pandey, Novel CFOA based capacitance multiplier and its application. AEU-Int. J. Electron. Commun. **107**, 192–198 (2019)
43. A. Yesil, E. Yuce, S. Minaei, Grounded capacitance multipliers based on active elements. AEU Int. J. Electron. Commun. **79**, 243–249 (2017)
44. E. Yuce, A novel floating simulation topology composed of only grounded passive components. Int. J. Electron. **97**(3), 249–262 (2010)
45. T. Yucehan, E. Yuce, A new grounded capacitance multiplier using a single ICFOA and a grounded capacitor. IEEE Trans. Circuits Syst. II Express Briefs **69**(3), 729–733 (2021)

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

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-022-20798-7

RESEARCH ARTICLE

A novel intelligent strategy–based thermodynamic modeling and analysis of solar-assisted vapor absorption refrigeration system

Harsh Kadyan[1] · **Anil Kumar Berwal**[1] · **Radhey Shyam Mishra**[2]

Received: 22 September 2021 / Accepted: 9 May 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

In recent times, solar energy has been utilized for refrigeration systems due to its efficiency and clean form of energy. Moreover, the evacuated tube collector (ETC)–assisted vapor absorption refrigeration system plays a significant role in the modern industrial world compared to the traditional electrical system. However, the conventional vapor absorption refrigeration system design is complex in nature and causes corrosion in the system. Therefore, in this research, novel Generalized Approximate Reasoning–Based Intelligent Control (GARIC) and Hybrid Ant Colony African Buffalo Optimization (HACABO) methods based on an ETC-linked 5-kW vapor absorption refrigeration system are proposed depending on lithium bromide-water (LiBr-H_2O). Primarily, the Haryana region's solar radiation and weather parameters were taken over a year to simulate the ETC system. ETC collects the solar energy for the refrigeration cycle, and the efficiency of the ETC is estimated using the GARIC method as per the input of solar radiation, collector area, and used solar energy. Moreover, the efficiency of the ETC is optimized using the proposed HACABO method. The modified polynomial fits curved equation is utilized for performance analysis. The simulation model of the solar cooling absorption system is carried out in the MATLAB platform. The coefficient of performance (COP) rate of the absorption cycle has gained 0.82% with the help of HACABO. Compared to other recent associated models, the proposed model has maximized the COP in the finest range.

Keywords Absorption cycle · Generalized approximate reasoning · Based intelligent control · Hybrid Ant Colony African Buffalo Optimization · Lithium bromide-water · Solar collector

Nomenclature

(S_a)	Solar power
G_s	General solar irradiance
τ_g	Transmittance of the glass cover
β_g	Absorption of the glass cover
c_t	Heating capacity
A_{sc}	Area of solar collector
$\hat{T}_{re}$	Average temperature
$\hat{T}_2$	Inlet of the water temperature
F_R^*	Heat removal factor
U_{hL}	Overall heat loss coefficient
g	Converted input
J	Exact consequent label
ϕ	Influence of efficiency
β	Influence of the desirability
$m_{i,j}$	Quantity of ETC efficiency
$b\hat{h}^*_{\max \cdot s}$	Data location
m_i'	Optimized efficiency
w_{i+1}'	New finest optimized efficiency
w_i'	Present efficiency
U	Total coefficient of heat transfer
$\hat{Q}_L$	Removed energy from the tank of storage
(kW)	Tank of storage
V_s	Volume

Responsible Editor: Philippe Garrigues

✉ Harsh Kadyan
harshkadyan@gmail.com

Anil Kumar Berwal
anilbeswal@gmail.com

[1] Centre of Excellence for Energy and Environmental Studies, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Sonepat, Haryana 131039, India

[2] Department of Mechanical Engineering, Delhi Technological University, Shahbad Daulatpur, Delhi 110042, India

Introduction

Solar energy has been utilized in a wide range instead of fuel uses; also, it has reduced several environmental threats made by vehicles or other electrical components (Salahuddin et al. 2018). In the past few decades, researchers have

Published online: 21 May 2022

made great advances in solar technology development and optimization mechanism of solar thermal collectors (Tabor et al. 2018), photovoltaic panels, and some hybrid modeled photovoltaic-thermal (PV/T) collectors, which are in the direct connection of sunlight (Kadyan et al. 2018). Solar collectors are gadgets designed to investigate thermal energy's solar radiation and transformation. Designing the cooling system using solar energy has attracted many researchers in recent years. Among the various solar energy–based thermal applications (Ebrahimnia-Bajestan et al. 2016), one of the most advanced constructions is the cooling system (Patel et al. 2009). Thus, the main feature of the solar cooling system model is solar absorption (Fong et al. 2010).

The absorption process holds and attracts moisture by its substances termed desiccants (Assilzadeh et al. 2005). Desiccants are also termed sorbents, which means the materials that can magnetize and handle other liquids or gases with a specified resemblance to water (Fekadu and Subudhi 2018). Moreover, the absorption schemes are the same as vapor compression (Julien et al. 2016) air condition modules but it varies in pressurization stages (Kalkan et al. 2012). To bear out the thermodynamic cooling strategy, the absorption cooling mechanism (refrigeration) can utilize a combination of two absorbent components and refrigerant. Moreover, the function of the solar-based air cooling system depends upon several components such as evacuated tube collector (ETC), absorber (here the LiBr absorbs the water from the atmosphere), heat exchanger, condenser, generator, evaporator, and expansion valve; also finally, it is connected to the required coolant surface.

Besides, the pressurization is gained by liquefying (Ker et al. 2012) the refrigerant in the absorber section. The heat generated in absorbent is utilized to specify the low-boiling in the refrigerant. In this manner, the compression happens without any need for huge mechanical energy (Kim and Infante Ferreira 2008). The absorption cooling model is the most frequently utilized mechanism of solar cooling. The vapor compression mechanism (Ventas et al. 2010) is also known as the absorption coolant machine, which realizes cooling by the expansion of fluid (Akbarzadeh and Wadowski 1996) refrigerant under condensed temperature and pressure, similar to the principle of conventional electrical processed vapor solidity air conditioner (Chidambaram et al. 2011). Normally two coolant mixtures are utilized: ammonia and water, and lithium bromide and water (Ferreira and Kim 2014). The ammonia and water–based vapor absorption system has attained less coefficient of performance (COP), and the conventional method tends to cause corrosion problems.

Also, the traditional method requires heavy components to enhance the system's performance, which increases the refrigeration system's initial cost. Most recently, particle swarm optimization (PSO) (Habibollahzade et al. 2021), evolutionary algorithm (Adham and Sulaiman 2017), genetic algorithm (Arshad et al. 2019), and multi-objective swarm optimization (Asadi et al. 2018) play a significant role to enrich the outcomes of ETC to increase the COP of the system. Moreover, artificial neural network (ANN) and fuzzy-based methods are conventionally utilized in the ETC extraction (Khosravi and Syri 2020). However, these methods do not perform well concerning the absorption cooling arrangement. Also, the working pressure of the conventional design is very less and this tends to diminish the COP very low. Thus, to enhance the solar cooling system model, this research focuses on solar-driven LiBr-H_2O absorption cooling arrangement (Kadyan and Berwal 2018), an intermittent system design in which water is utilized for the cooling system; also, LiBr is used as an absorber. The LiBr-H_2O absorption model with a thermal capability of 5 kW is to be verified. However, the finest performance is necessary to design a 5-kW vapor absorption refrigeration system.

Therefore, this research focuses on the study of an effective 5-kW vapor absorption refrigeration design employing significant components and innovative approaches. The solar ETC is utilized to gather the heat from the solar. Several ETC models such as matching-based control (Han et al. 2019), desiccant-coated heat exchanger (Xu et al. 2019), nanofluid for absorption (Nourafkan et al. 2019), and gradient optimization ETC (Pang et al. 2019) have been introduced in the past, but still suitable outcomes are not attained.

Thus, the conventional ETC is not enough to gain the desired energy efficiency. So, a novel GARIC-based HACABO has been introduced for the vapor absorption system. The 5 kW has been gained by optimizing the ETC using the hybrid optimization model. Moreover, the present energy efficiency in the ETC is predicted using the GARIC model. Moreover, the predicted efficiency is optimized using a novel hybrid form of the HACABO-based optimization method. In addition, this experiment is done in the Haryana region at different weather conditions. Finally, key parameters like COP and thermal efficiency have been optimized in the MATLAB environment. Furthermore, the improvement score has been found by making a comparative analysis. The values of COP and heat have verified validation of this method, and it is compared with the different existing solar-linked cooling systems.

The organization of the paper is summarized as follows. "Related work" describes a small review related to the part of the work. "System model and problem statement" explains the system description and problem. "Proposed GARIC-HACABO approach" describes the proposed GARIC regression, hybrid optimization, and 5-kW absorption system design. "Results and discussion" contains the results and discussion. Eventually, the conclusion is described in "Conclusion."

Related work

Some of the recent literature related to solar-based cooling systems is summarized as follows:

People often live under two weather conditions: chill and strong heat, as of global environmental troubles. Thus, continuous enhancement in living standards, air cooler, refrigeration system, and air conditioning has become broadly used worldwide. So Han et al. (2019) proposed an impedance matching–based control strategy to regulate the cooling system by a solar distributed system. Hence, the experimental work achieved a photoelectric conversion measure reaching 0.29 approximately and enhanced by 83.7%, which is close to the control system but without the use of a controller. However, effective cooling is not possible with this approach because the attained COP value is 0.62%.

To offer fulfilled air is provided with sufficient humidity and temperature ratio in the system. However, the conventional vapor firmness cooling method has to dehumidify the air via reducing its temperature to less than its actual amount and reheating the air to reach the deliver state of temperature that tends to increase the energy dissipation. To overcome this problem, Xu et al. (2019) proposed a desiccant-coated heat exchanger-based cooling scheme model. Thus, the desiccant system is fabricated in high priority metal-organic framework. At last, the achieved result proved the efficiency of the proposed cooling system with a COP rate of 0.71%. Nevertheless, the design of this model requires more energy and is complex in nature.

Normally, for the solar cooling system, LiBr-H_2O or ammonia-water is used. Moreover, the concentration of LiBr salt varies from one model to another; also, part of the refrigerant system is varied by its rate of range and weight. For this purpose, a binary nanofluid for absorption in the solar refrigeration system was introduced by Nourafkan et al. (2019). The experimental results verified that the solar radiation enhanced up to 4.9% and achieved high stability. The COP of this approach is less than of other methods.

Nowadays, solar technology plays a great role in the digital world in the energy consumption of many applications. However, there are some difficulties in cooling load with an insufficient collector array. For this reason, Ye et al. (2019) proposed a sub-coolant-based absorbent module in the cooling system. Moreover, it is one of the new ideas in the solar cooling system model; thus, the proposed model has improved the system performance by its temperature and flow rate. But, the design of this system is complex in nature with high cost.

With the provocation of the comprehensive energy disaster and ecological effluence, electrical vehicles are urbanized to receive the position of the conventional vehicles gradually, because of their reduction emission and energy-savings. So Pang et al. (2019) proposed a solar-linked air cooling method and gradient optimization for vehicles. The development of the solar-based cooling system in vehicles is to reduce air pollution; thus, the proposed model was implemented successfully and reduced air pollution. However, the appropriate enhanced technique is not validated in this cooling model.

Bhowmik et al. (2021) developed an adaptive neuro-fuzzy inference system (ANFIS)–based method for the system improvement in solar ETC–based liquid desiccant cooling application. The validation has been developed with diverse desiccant flow values. Consequently, the average absolute error is achieved as 2.3%. However, the effective kW design is not validated in this work.

The review detailed that the conventional method has more complexity to design the cooling application. Consequently, the value of COP is very less for effective performance, and also, the cost is high. The exact cooling capacity design was not possible with the traditional refrigeration system, and the efficiency of the appropriate value was not tracked using the conventional design of the refrigeration system. Hence, this research developed a new kW-based cooling system design based on the improved intelligent strategy and optimization approach.

The main contributions of the research are summarized below:

- Initially, design the ETC model for generating heat energy from the solar.
- Develop LiBr-H_2O-based 5-kW vapor absorption refrigeration model using MATLAB tool.
- For the valuation of solar radiation, the Haryana region's solar and weather data is has been gathered in a real-time basis.
- Moreover, a novel GARIC approach is proposed for finding the finest efficiency from the ETC model.
- Also, the efficiency from the ETC is optimized by the novel HACABO technique.
- The performance analysis is done by modified polynomial fits curved equations.
- Finally, the performance of the proposed cooling system design is compared with various conventional methods in terms of COP and thermal efficiency.

System model and problem statement

In the vapor absorption refrigeration system, LiBr-H_2O is taken as an absorbent. Here, LiBr is considered the solution, and H_2O acts as the refrigerant. This vapor absorption refrigeration scheme functions only above zero degree temperature. The absorption refrigerant method is mainly based on the quantity of heat supply from outside, internal heat, and the heat exchange rate between the LiBr and water. Moreover, the water collects the heat from the ETC,

and it is impelled to the storage tank; then, the LiBr-H_2O is heated through the generator. A low vapor and temperature are attained in the evaporator (point 10). The H_2O in the form of water is passed into the absorber, and it appears with the strong LiBr solution. The schematic illustration of the solar-assisted vapor absorption refrigeration method is illustrated in Fig. 1. Moreover, the water vapor from the generator is also absorbed, and this process forms a weak LiBr solution. Furthermore, the strong LiBr-H_2O solution from the generator system is sprayed as a fine droplet in the absorber. In the absorber, the heat $\hat{Q}_A$ is discarded into the surroundings for maintaining the pressure and temperature using the cold water flow. The low-pressure weak solution is forced to the generator using a pump via the heat exchanger.

Consequently, the strong solution from the pump via the heat exchanger unit has increased the temperature of the strong solution. The generated heat from the generator produces the temperature to the solution converted into vapor, which flows to the condenser. Then, the condenser condenses and converts the vapor into a pure liquid of H_2O. Nevertheless, the high-pressure H_2O from the condenser flows into the expansion valve and is converted into low pressure; this low-pressure water is sprayed in the evaporator (point 9). While the water entering the evaporator gets cooled by the coil, the refrigerating system effect is taken as $\hat{Q}_E$. Furthermore, from the evaporator, the water is converted into vapor which flows to the absorber, and this process of vapor absorption refrigeration cycle is terminated. Besides, the high LiBr salt from the generator converts into a strong solution. This solution flows into a heat exchanger under high pressure. The heat exchanger turns the solution into a weak solution; therefore, the heat is reduced. The solution at low pressure is again sprayed into the absorber and this cyclic process is terminated. However, the effective energy intriguing from solar is highly crucible. Moreover, during the absorbent operation, the high concentration of LiBr tends to create a crystal that can chunk the pipes. These will affect the COP of the refrigeration system. Therefore, a proper design is necessary to enhance the performance of the system.

Proposed GARIC-HACABO approach

The proposed GARIC-HACABO-based 5-kw LiBr-H_2O-based vapor absorption system is illustrated in Fig. 2. The proposed strategies investigate the solar radiation assessment of the Haryana region. This research focuses on the model of

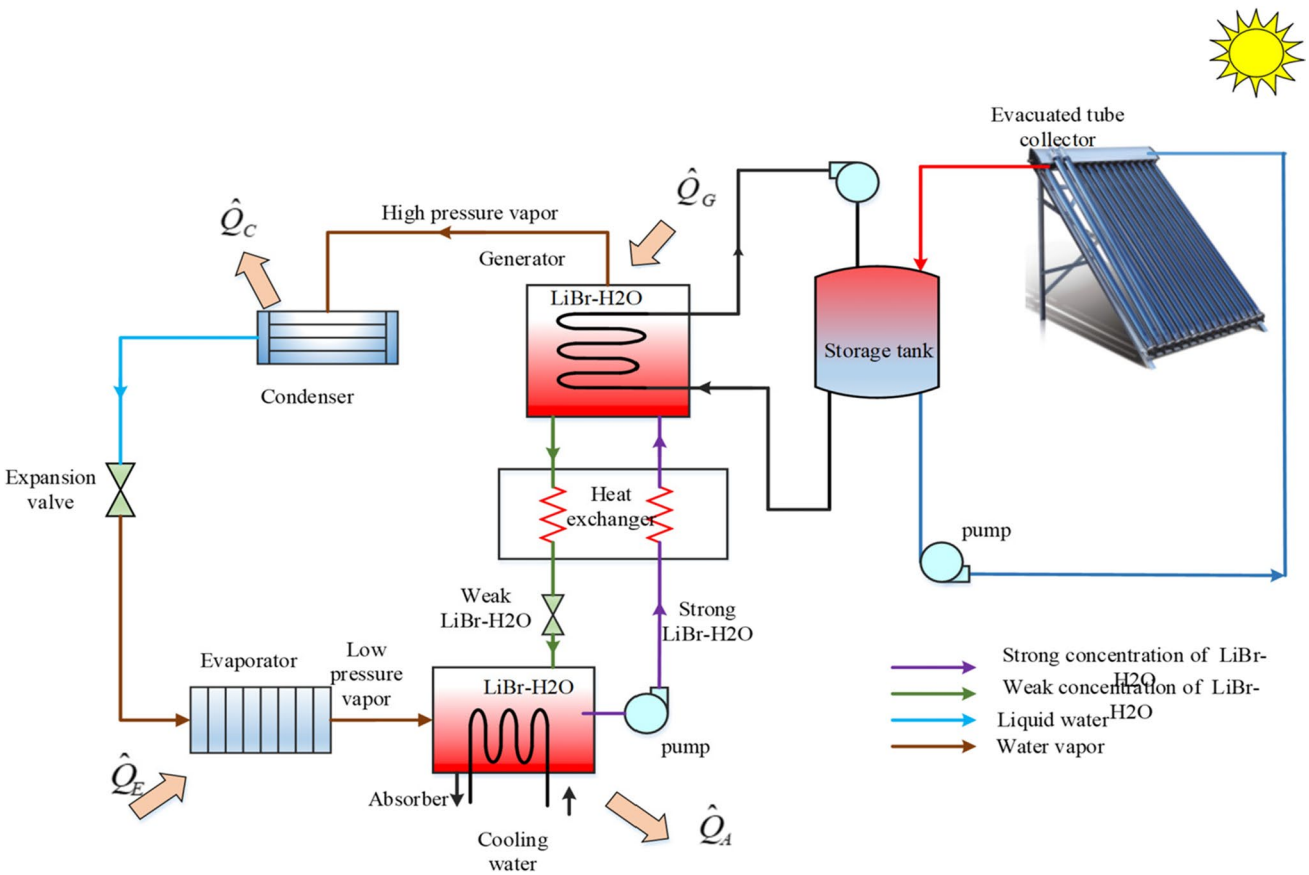


Fig. 1 System model of solar-assisted LiBr-H_2O-based vapor absorption refrigeration system

the solar-based cooling system with a 5-kW capacity vapor absorption system. The radiation of the solar energy is collected by the ETC.

Besides, the finest efficiency of the ETC is tracked by the novel GARIC approach. The parameters of ETC efficiency are optimized using the novel HACABO method. Finally, the performance analysis is done using a modified polynomial fits curved equation. The absorption scheme model is made by LiBr-H_2O that can absorb 5kW vapor. Besides, based on some specifications adiabatic absorber with the vapor absorption refrigeration method is analyzed such as at the evaporator the cooling capacity $\hat{Q}_E$ is 5kW, the condenser temperature is considered as $\hat{T}_C$=40 °C for peripheral descend, employing the refrigerant the evaporator temperature is taken as at $\hat{T}_E$=40 °C, the absorber temperature is assigned in the range of 25–60 °C, and also the generator temperature is assigned in the range of 65 to 90 °C. Moreover, the mass concentration of LiBr and the particular enthalpy of the solutions were evaluated by associations available from the handbook of ASHRAE.

Modeling of ETC

The solar ETC is a series connection of evacuated tubes for heating water. When the solar energy is captured in the vacuum space of the tube, the heat loss to the surroundings diminishes (Zambolin and Col 2010). The internal metal tube of the ETC acts as the absorber plate, and it is linked with the heat pipe to take the collected heat from the sun to water. The solution inside the heat pipe is under extremely meticulous pressure. In this condition, the boiling liquid is placed at the end of the hot side of the pipe although the condensing vapor occurs at the cold end. Consequently, it permits the transfer of thermal energy from one side of the pipe to another. Moreover, the heat from the sun transfers from the hot end to the cold end; then, the thermal energy is transmitted into the water for the use of heating. Evacuated tube collectors commonly have greater effectiveness over a complete year as they may still manage under cloudy conditions. The occupied solar power (S_a) is identical to the proceeding solar power time's optical disappearance and is demonstrated in Eq. (1):

$$S_a = \left(\tau_g \beta_g\right) G_s A_{ca} \tag{1}$$

where β_g is the absorption of the glass cover and τ_g is the transmittance of the glass cover, A_{ca} is the collector opening area, and G_s is the general solar irradiance on the collector surface. Then, the performance of the collector is based on the first law of thermodynamics energy balance, which means the energy utilized is identical to the energy absorbed:

$$\sum \hat{E}_{in} = \sum \hat{E}_{out} \tag{2}$$

The amount of energy absorbed from the solar collector is equivalent to the amount of energy used to increase the operating temperature of the fluid in Eq. (2). Consequently, the energy utilization can be estimated by Eq. (3):

$$S_a = \tilde{m} c_t \left(\hat{T}_2 - \hat{T}_1\right) \tag{3}$$

where the water mass flow rate as working fluid is represented as $\tilde{m}$, the water heating capacity is referred as c_t, the inlet of water temperature is denoted $\hat{T}_2$, and $\hat{T}_1$ is the outlet of water temperature. Although, the amount of energy absorbed by the ETC is evaluated in Eq. (4) as

$$S_a = A_{sc} \eta_F \left[G_s \left(\tau_g \beta_g\right) - U_{hL} \left(\hat{T}_{re} - \hat{T}_a\right)\right] \tag{4}$$

where the area of the solar collector is A_{sc}, efficiency factor is denoted as η_F, the solar radiation is G_s, coefficient of heat loss is U_{hL}, the transmittance, as well as absorptance, is represented as $(\tau_g \beta_g)$, the refrigerant average temperature is $\hat{T}_{re}$, and the ambient temperature is $\hat{T}_a$, respectively. The

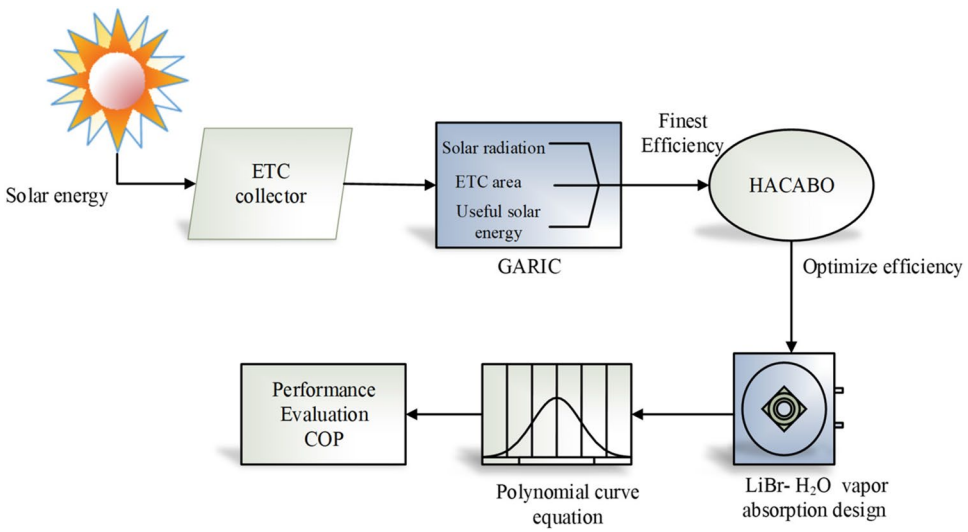


Fig. 2 The proposed GARIC-HACABO-based 5-kw LiBr-H_2O-based vapor absorption system

coefficient of heat loss in the absorber tube is estimated using Eq. (5) as

$$U_{hL} = \frac{\tilde{mc}_{sh}\left(\hat{T}_1 - \hat{T}_2\right)_{av}}{A_{sa}\left(\hat{T}_{mw} - \hat{T}_{at}\right)_{av}} \tag{5}$$

where c_{sh} is the exact water heat, the outlet and inlet of the water temperature are referred as $\hat{T}_2$ and $\hat{T}_1$, the absorber surface area is denoted as A_{sa}, and the average dissimilarity among the average temperature of the water and the ambient temperature than the epoch time is $\left(\hat{T}_{mw} - \hat{T}_{at}\right)_{av}$. The efficiency of the ETC is expressed in Eq. (6) as

$$\eta = \frac{S_a}{A_{sc}G_s} \tag{6}$$

$$\eta = F_R^*\left(\tau_g\beta_g\right) - F_R^*U_{hL}\left(\frac{T_{in}^* - T_a^*}{G_s}\right) \tag{7}$$

where F_R^* is the heat removal factor and U_{hL} is the overall heat loss coefficient, which is demonstrated from Eqs. (4) to (6). The solar thermal energy has been collected by the ETC model and the utmost efficiency from the ETC is obtained using the GARIC and HACABO methods for improving COP of the refrigeration system.

Proposed GARIC for ETC efficiency

The proposed GARIC is a new neuro-fuzzy augmentation learning form for the estimation of ETC efficiency. Here, the action selection network (ASN)–based GARIC method is used for this work. The ASN-based GARIC method has five layers of network with each layer working in one phase of the fuzzy inference method. The GARIC architecture of ASN in ETC efficiency is shown in Fig. 3.

The input of the first layer accepts the solar radiation, ETC area, and solar power as the real-valued variables. Here, the crispy input is converter into the fuzzy input by means of the fuzzification process. Then, the second layer is the forbear rule layer, and at the second layer, the input of the first one is provided which works based on the rules function. The triangular membership function is applied in the second layer based on the linguistic variables from the earlier layer that is estimated using Eq. (8) as

$$\lambda(g) = \frac{1}{1 + \left|\frac{g - c'}{e'}\right|^n} \tag{8}$$

where $e' = e'_L$ or e'_R represents the left or right expansion of the triangular membership functions and g is the converted input. Also, the condition $g < c'$ or $g \geq c'$ is articulated and the curvature controlled function is represented as n. The function of the triangular membership rules for the second layer is estimated by Eq. (9) as

$$\lambda_{c',e'}(g) = \begin{cases} 1 - |g - c'|/e'_R & g \in \left[c' + e'_R\right] \\ 1 - |g - c'|/e'_L & g \in \left[c' - e'_L\right] \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

The expansion and the center are the weights of the gathered data in the membership function considered the input links. Afterwards, the third layer is connected with the forebear rule conditions, which participate in the

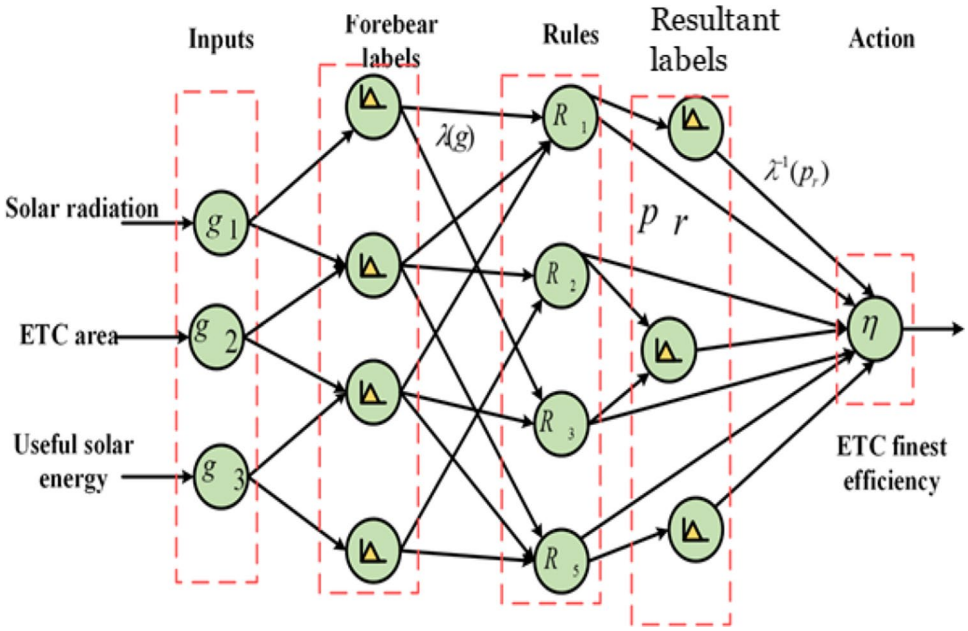


Fig. 3 GARIC architecture for ETC efficiency

second layer condition part of that rule. The minimum action of the third layer is performed by itself which is expressed by Eq. (10):

$$p_r = \frac{\sum_k \lambda_k e^{-r\lambda_k}}{\sum_k e^{-r\lambda_k}} \tag{10}$$

Here λ_k is the degree of the forbear rule and the resultant of the input variable. The minimum operation rigidity is maintained by the constraint r. Moreover, at the instance of the learning process, the gradients are calculated in finite r. In the fourth layer, all the rules from the input are given, which work as a consequent label. Based on the rule, this nodule is suggested to the action of output. The mapping of this action is evaluated by Eq. (11):

$$\lambda^{-1}_{c'J,,e'_{JL},e'_{JR}}\left(p_r\right) \tag{11}$$

where J is the exact consequent label; if the purpose is monotonic, then the λ^{-1} will not exist. The local mean of maximum triangular function is given in Eq. (12) as

$$\lambda^{-1}_{c'J,,e'_{JL},e'_{JR}}\left(p_r\right) = c'_J + \frac{1}{2}\left(e'_{JR} - e'_{JL}\right)\left(e - p_r\right) \tag{12}$$

In this case where $p_r = 0$, the $c'_J + \left(e'_{JR} - e'_{JL}\right)/2$ is obtained by the limiting value of $\lambda^{-1}(p_r \to 0)$. From this, the membership function of the triangular median is projected as a set of $\lambda^{-1}(0, 1)$. In the next layer, the computation process is taken and in the fifth layer, the action of output variables attained the finest ETC efficiency. The output produces the action defuzzification rule method of execution, which is evaluated using Eq. (13):

$$\eta = \frac{\sum_{r=1}^{n} p_r \lambda^{-1}\left(p_r\right)}{\sum_{r=1}^{n} p_r} \tag{13}$$

The action η is done which is expressed by the forebear input label functions. From the output, it shows the exact efficiency of the ETC.

HACABO for parameter optimization

ACO and ABO are combined as the HACABO method based on their probabilistic and better communication techniques. The flowchart of HACABO for ETC efficiency parameter optimization is shown in Fig. 4. In ACO, pheromone trails are considered fitness and the exploitation and exploration in ABO are taken as fitness. Initially, the ETC parameters and the obtained finest efficiencies are initialized in the HACABO algorithm. The initial selection process of the construction probability stage is expressed by Eq. (14):

$$\widehat{P}_{i,j} = \frac{\left(m^{\phi}_{i,j}\right)\left(w^{\beta}_{i,j}\right)}{\sum\left(m^{\varphi}_{i,j}\right)\left(w^{\beta}_{i,j}\right)} \tag{14}$$

where $m_{i,j}$ is the quantity of ETC efficiency of i, j data, ϕ is the influence of efficiency, $w_{i,j}$ points out the desirability of i, j data, and β is the influence of the desirability, respectively. Moreover, the fitness of ABO is added in place of the ACO fitness stage for enhancing the fitness function of both algorithms. The movement of finest ETC efficiency is optimized using Eq. (15):

$$m'_{i+1} = m'_i + lc1\left(b\hat{g}^*_{\max} - m'_v\right) + lc2\left(b\hat{h}^*_{\max \cdot i} - m'_i\right) \tag{15}$$

where m'_i is the movement of the efficiency at the particular location of ETC at $(i = 1, 2, .. \ldots n)$, the finest ETC efficiency in the data location is referred as $b\hat{g}^*_{\max}$, the finest data location created by a particular ETC efficiency is referred as $b\hat{h}^*_{\max \cdot s}$, and the parameters for learning are denoted as $lc1$ and $lc2$, which belongs to [0, 1].

Consequently, the movement of present optimized efficiency m'_i of the ETC to the new finest optimized efficiency is identified as w'_{i+1}, which is evaluated by Eq. (16):

$$w'_{i+1} = \frac{\left(m'_i + w'_i\right)}{\gamma} \tag{16}$$

where the new optimized efficiency of the ETC movement is referred as m'_{i+1}, and the exploration range of present efficiency is referred as w'_i. Once the best output of optimized efficiency is obtained, the process is stopped until it repeats the process from the initial level until the finest outcome attains.

Proposed absorption cycle modeling

The LiBr-H_2O absorption cooling scheme model is established based on the laws of mass and also energy protection by acquiring total power across each of the parts, namely storage tank, evaporator, heat exchanger, absorber, condenser, and generator.

LiBr-H_2O is used for moderated temperature refrigeration systems. Here, the LiBr solution is utilized as an absorbent, and water is used as a refrigerant. The solution concentration in the absorption chamber is referred to as the ratio of Li-Br mass to the sum of solution mass, which is expressed in Eq. (17).

$$C = \frac{\widetilde{m}_{LiBr}}{\widetilde{m}_{LiBr} + \widetilde{m}_{H_2O}} \tag{17}$$

The LiBr-H_2O solution of equilibrium temperature is attained via Eq. (18) as

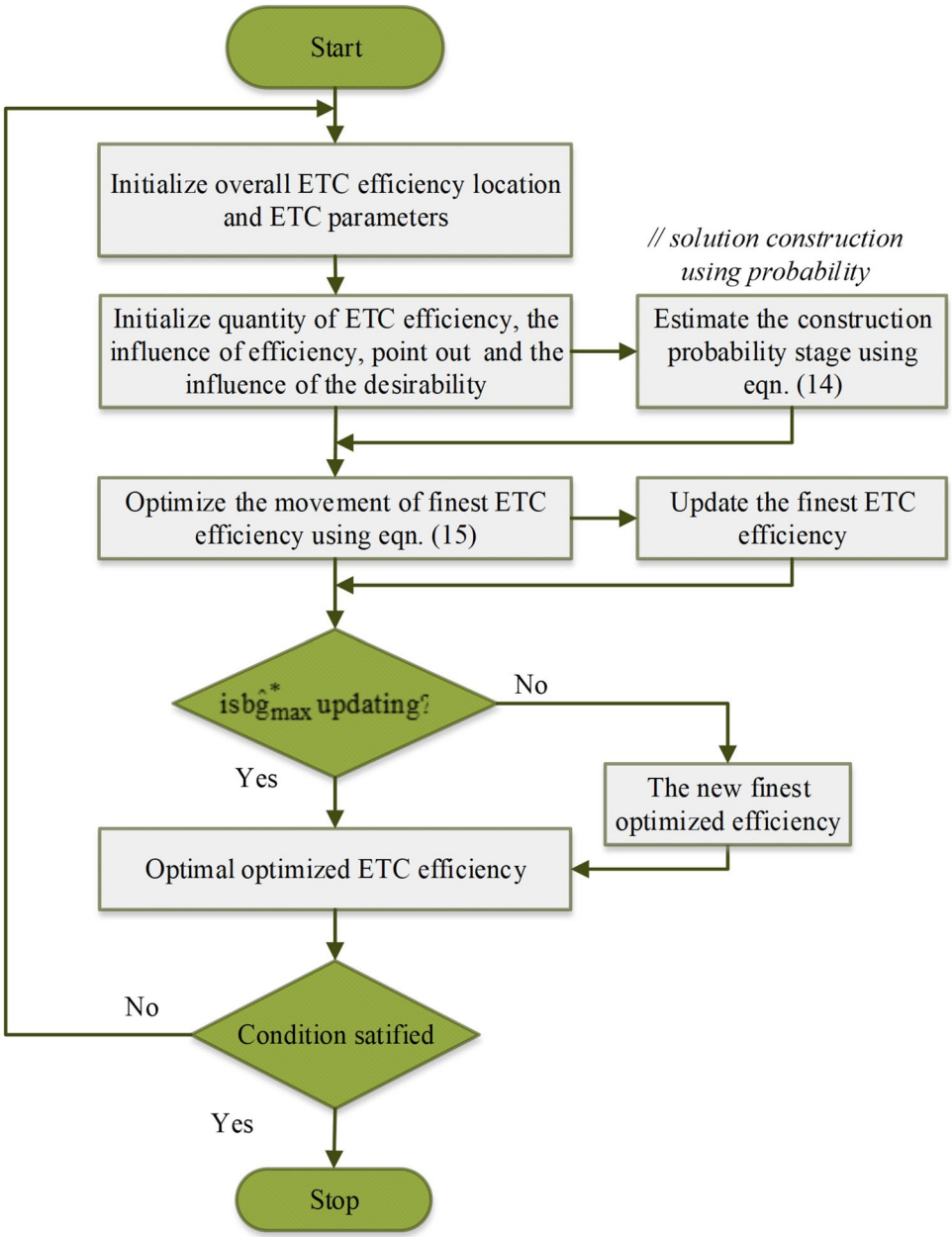


Fig. 4 Flowchart of HACABO for optimizing the ETC efficiency

$$T_s = T_{rf} \sum_{k=0}^{3} a_k C^k + \sum_{k=0}^{3} b_k C^k \tag{18}$$

The LiBr-water solution of equilibrium enthalpy can be attained by Eq. (19) as in the modified polynomial curve fitting approach:

$$h_s = \sum_{k=0}^{4} c_k C^k + T_s \sum_{k=0}^{4} d_k C^k + T_s^2 \sum_{k=0}^{4} e_k C^k \tag{19}$$

where, $a_k . , \ldots \ldots e_k$ represent the constant coefficients and h_s is the polynomial curve fitted enthalpy. The error value is estimated from the polynomial derivation, which is defined as using Eq. (20):

$$e = \sum_{k=0}^{n} \left(h_y - h_s\right)^2 \tag{20}$$

where h_y is the similar point of h_s and n is the amount of points in enthalpy.

In the storage tank, the tank head portion lead conducts maximum temperature and the base portion conducts the lowest temperature. This obtained energy is based on the capacity of the tank, mass of the fluid entry, and exit. The

temperature for the storage tank outlet water is expressed in Eq. (21):

$$\hat{T}_{s,\text{new}} = \hat{T}_{s,\text{old}} + \frac{\Delta t'}{\tilde{m}.\left(C_v^* w\right)s} \cdot \left[S_a - \hat{Q}_L - (U.A)_s - \left(\hat{T}_s - \hat{T}_0\right)\right] \tag{21}$$

where S_a is the useful collected energy (KJ), $\hat{Q}_L$ is the removed energy from the tank of storage (kW), U is the total coefficient of heat transfer, C_v^* is the specific heat at constant volume, and $\hat{T}_s$ is the storage temperature. The thermal capacity of the thermal storage is in Eq. (22):

$$\hat{Q}_S = \sum \left(L_i' - \hat{Q}_0\right).\Delta t_i' \tag{22}$$

In the generator unit, the function of the generator is to provide the vapor of refrigerant to the system. Moreover, the solution of the refrigerant is separated in the system. The heat of warm water is absorbed by the solution and split the solution of absorbent from water. The solutions of concentrated absorbent flow back to the absorber and the unit of condenser refrigerant vapor. Consequently, the heat rate is given to the generator, which is the rate of energy input to the cycle that can be obtained from the balanced energy in Eq. (24) as follows:

$$\tilde{m}_3 = \tilde{m}_4 + \tilde{m}_7 \tag{23}$$

$$\hat{Q}_G = \tilde{m}_7 h_7 + \tilde{m}_4 h_4 - \tilde{m}_3 h_3 = \tilde{m}_r h_7 + \tilde{m}_s s h_4 - \tilde{m}_w s h_3 \tag{24}$$

where $\tilde{m}_w s = \tilde{m}_r + \tilde{m}_s s$, the strong solution mass flow rate is denoted as $\tilde{m}_s s$, the weak solution mass flow rate is termed as $\tilde{m}_w s$, and the mass flow value of refrigerant is denoted as $\tilde{m}_r$. The solution in the absorber absorbs the refrigerant vapor. From this, vapor in the condenser is converted into a liquid and the created heat is attained in the evaporator. Then, the heat removal rate from the absorber is given in Eq. (25):

$$\hat{Q}_A = \tilde{m}_{10} h_{10} + \tilde{m}_6 h_6 - \tilde{m}_1 h_1 = \tilde{m}_r h_{10} + \tilde{m}_s s h_6 - \tilde{m}_w s h_1 \tag{25}$$

The total mass balance is evaluated by Eq. (26).

$$\delta = \tilde{m}_{ss} / \tilde{m}_r \tag{26}$$

Hence

$$\tilde{m}_{ss} = (1 + \delta)\tilde{m}_r \tag{27}$$

The pure water mass balance is estimated using Eq. (28):

$$\tilde{m}_r + \left(1 + \kappa_{ss}\right)\tilde{m}_{ss} = \left(1 + \kappa_{ws}\right)\tilde{m}_r \tag{28}$$

$$\delta = \kappa_{ws} / \left(\kappa_{ss} - \kappa_{ws}\right) \tag{29}$$

Therefore

$$\hat{Q}_A = \tilde{m}_r h_{10} + \delta \tilde{m}_r h_6 - (1 + \delta)\tilde{m}_r h_1 \tag{30}$$

In this unit, the function of the condenser is to reduce the vapor of refrigerant and tubes, the cooling water pours, and then, the hot vapor of refrigerant fills the surrounding space. Furthermore, the bottom of the condenser is a liquid refrigerant that is stored before the drive to the expansion valve. The heat rejection rate of the condenser is determined by Eq. (32):

$$\tilde{m}_7 = \tilde{m}_8 = \tilde{m}_r \tag{31}$$

$$\hat{Q}_C = \tilde{m}_7\left(h_7 - h_8\right) = \tilde{m}_r\left(h_7 - h_8\right) \tag{32}$$

In this unit, the function of the evaporator is to chill the surrounding water. Moreover, the heat is absorbed by the refrigerant in the surrounding water; then, it evaporates. This leads to an increase in the pressure of the vessel. Thus, the heat rate added to the evaporator is the cooling effect which is created by the absorption cooling system and it is determined via Eq. (33):

$$\tilde{m}_8 = \tilde{m}_9 = \tilde{m}_r$$

$$\hat{Q}_E = \tilde{m}_9\left(h_9 - h_8\right) = \tilde{m}_r\left(h_9 - h_8\right) \tag{33}$$

The energy equilibrium on the hot along with the cold side of the solution heat exchanger is defined using Eq. (36):

$$\tilde{m}_3 = \tilde{m}_2 = \tilde{m}_{ws} \tag{34}$$

$$\tilde{m}_5 = \tilde{m}_4 = \tilde{m}_{ss} \tag{35}$$

$$\hat{Q}_{HE} = (1 + \delta)\tilde{m}_r\left(h_3 - h_2\right) = \delta \tilde{m}_r\left(h_4 - h_5\right) \tag{36}$$

The liquid form of refrigerant is moved into the evaporator via expansion. Consequently, the dissimilarity among the high pressure along with the low pressure of condenser and evaporator is maintained here. The mixture of vapor and liquid can move into the evaporator. The expansion valve mass balance is defined in Eq. (37):

$$\tilde{m}_9 = \tilde{m}_8 = \tilde{m}_r \tag{37}$$

$$h_8 = h_9 \tag{38}$$

The solution pump is defined through Eq. (39) as follows:

$$\tilde{m}_1 = \tilde{m}_2 = \tilde{m}_w s \hat{W}_p = (1 + \delta)\tilde{m}_r V_s\left(P_C - P_E\right) \tag{39}$$

where the particular solution volume is represented as V_s. The COP of the absorption system is determined as the ratio

of heat absorbed by the absorber to the heat rate input to the generator as defined in Eq. (40):

$$\mathrm{COP} = \frac{\hat{Q}_E}{\hat{Q}_G + \hat{W}_p} = \frac{\hat{Q}_E}{\hat{Q}_G} \tag{40}$$

Moreover, the efficiency of the second law approach is evaluated by Eq. (41) as

$$\eta_S = \frac{\mathrm{COP}}{\mathrm{COP}_{\max}} = \frac{\hat{Q}_E}{\hat{Q}_G}\left(\frac{T_G}{T_G - T_C}\right)\left(\frac{T_G - T_E}{T_E}\right) \tag{41}$$

Based on this modeling, the effective 5-kW vapor absorption refrigeration system has been developed with the effective performance.

Results and discussion

In this section, the proposed method of solar-assisted 5kW-based vapor absorption refrigeration system is executed in the MATLAB/Simulink R2018b environment and the platform of Windows 7. It is a mathematical and restrictive programming language formed by MathWorks. The implementation model was utilized for the evaluation of LiBr-H_2O-based vapor absorption refrigeration method design constraints. Furthermore, the contour graphs and other related graphs were designed by the constraints. Primarily, the optimal efficiency of ETC is gathered using GARIC and HACABO methods. Then, the optimized efficiency is provided to the absorption cycling model. The program flow chart of the LiBr-H_2O absorption scheme is demonstrated in Fig. 5.

The constant constraints of the designed vapor absorption system are 228.9 m^3/kg liquid, 181.6 m^3/kg latent, 410.5 m^3/kg vapor, and pressure 51.14 kPa(abs). Moreover, the Haryana air condition data was collected in the real-time basis; then, the vapor absorption system performance has been noted in each weather condition. To get the optimized ETC parameter results, software simulation is done in the MATLAB environment. Hence, to train the data in the MATLAB system, the GARIC neural approach is used.

- Initially, the input parameters $\hat{T}_G$, $\hat{T}_A$, $\hat{T}_E$, $\hat{T}_C$, $\tilde{m}_r$, and $\hat{Q}_E$ were taken for the design and the constraints are estimated using observed equation, catalog of enthalpies, and LiBr-H_2O strong as well as weak solutions.
- Furthermore, based on the enthalpies, mass flow rate, and $\hat{Q}_C$, the parameters for design are taken.
- If the LiBr-H_2O weak solution is superior to the LiBr-H_2O strong solution, then the second set of design parameters is evaluated using $\hat{Q}_G$ and remaining mass flow rates.
- Then, the COP of the design is estimated and all unknown values are evaluated by the operating area.
- If the LiBr-H_2O weak solution is lesser than the LiBr-H_2O strong solution, then all unknown values are estimated at a non-operating area.
- Finally, the results are plotted by the representation.

Case study

Consider the plant includes 5-kW cooling absorption system, evacuated tube collector area of 28.20m^2. The efficiency of the collector field is 0.96 under the temperature 90 to 150 °C, and the water flow rate is 0.02g/h; when the solar irradiance and ambient temperature are 700 W/m^2 and 20°C, the total energy generated is 2 kWh. The parameters of ETC are exposed in Table 1.

The production of solar energy is normally higher than the requirement of the generator, which comes from the tilt angle of the ETC, and the design standards are used. In this ETC method, the utmost ambient temperature is chosen for the standard design but the seasonable solar radiation is absent at that time. Therefore, the collector generates more energy than is necessary and at this point, the solar radiation power (S_a) is greater than700kW. The solar radiation $G_s$of 700W/m^2, ETC area A_{ca} of 28.20m^2, and solar power (S_a) of 700kW are given as the input of the GARIC method. A certain triangular membership function is applied to the system for finding better efficiency. The limiting rule has been applied in all the layers of the GARIC method and the finest efficiencies are tracked finally using Eq. (13) and achieved value as follows:

$$\eta = \frac{\sum_{r=1}^{n} 28.20 \times 700^{-1}(28.20)}{\sum_{r=1}^{n} 28.20} = 80\%$$

Then, the tracked ETC efficiencies are provided to the HACABO method. The obtained efficiency is optimized using the HACABO approach. The optimized ETC efficiency is achieved using Eq. (15) as m'_{i+1}=96%. The predicted and optimized ETC efficiency using GARIC-HACABO methods is represented in Fig. 6.

The optimal validation of the solar ETC efficiency using the GARIC and HACABO methods is shown in Fig. 7. Moreover, the energy has been saved by the hot storage and the storage capacity of the hot water storage tank is considered 1.3m^3. The water flow is passed to the ETC using a speed pump and this improves the working pressure of the system.

The difference in temperature between the inlet and outlet of the ETC is influenced by the energy amount distributed to the generator. The rate of flow to the collector is required to be minimizing because the energy amount to the generator is constant by increasing the temperature at diverse conditions. The LiBr solution physical parameters used in the simulation

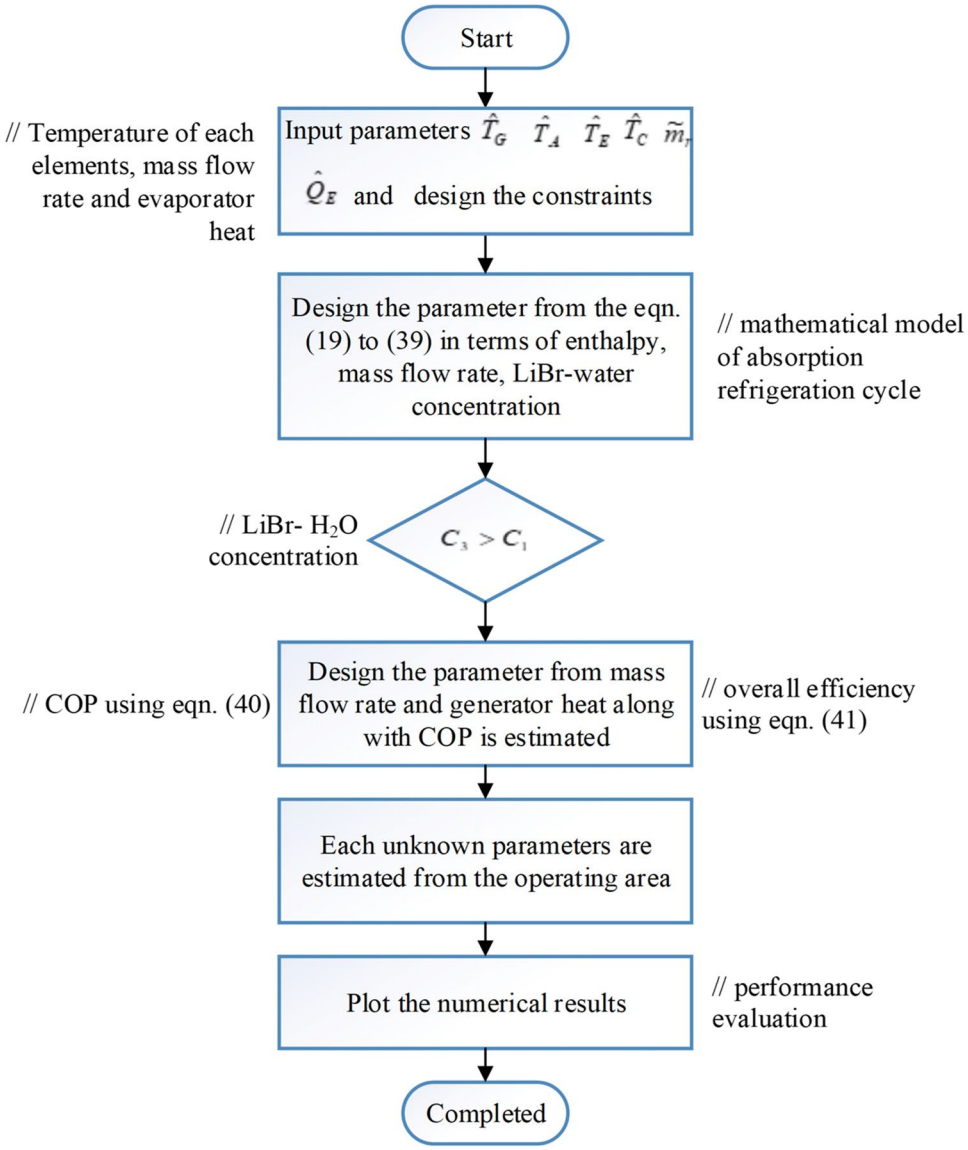


Fig. 5 Program flow chart of a LiBr/water absorption system

purpose are estimated which include specific heat capacity, enthalpy density, and vapor pressure. The relation between the concentration of LiBr solution and the element's temperature is approximately linear. The process working in the pressure-enthalpy representation of LiBr-H_2O is illustrated in Fig. 8. To

Table 1 The evacuated tube collector characteristics

Description	Parameter	Ranges
Inlet temperature of the collector	$\hat{T}_2$	50–200°C
Outlet temperature of the collector	$\hat{T}_1$	40–95°C
Capacity of hot water inside the collector	c_t	53–85°C
Ambient temperature	$\hat{T}_a$	12–33°C
Flow rate of hot water	$\tilde{m}$	7–10m^3/h
Solar radiation	G_s	200–900W/m^2
Efficiency of ETC	η	60–96%

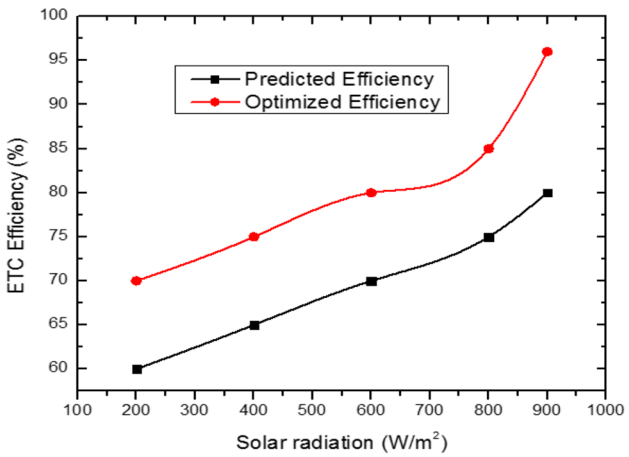


Fig. 6 The predicted and optimized ETC efficiency using GARIC-HACABO methods

analyze the changing condition of the refrigeration system in different states, that is, liquid to gas or gas to liquid, the mass flow rate has been measured in the form of pressure-enthalpy in Fig. 8. It is the mass flow percentage rate of the proposed system.

The LiBr-H_2O enthalpy, mass flow rate, pressure, temperature, and concentration are shown in Table 2. The LiBr-water of the vapor absorption coefficient of performance under the generator temperature $\hat{T}_G$ is 80°C, and absorber temperature $\hat{T}_A$ is 25°C, when the condenser temperature $\hat{T}_C$ being 30°C and the evaporator temperature $\hat{T}_E$ being 7°C are fixed respectively.

Mathematical calculation of each component in the vapor absorption refrigeration cycle:

$$\tilde{m}_3 = 0.02785 + 0.00275 = 0.0306\ kJ/kg^{-1}$$

The heat removal rate from the absorber is

$$\hat{Q}_A = 0.00275 \times 2606.65 + 0.02785 \times 1760.0274 \times 168 = 5.0071\text{kW}$$

Applying the energy balance: $\tilde{m}_w s = 0.0274\ kJ/kg^{-1}$, $\tilde{m}_r = 0.0075\ kJ/kg^{-1}$, and $\tilde{m}_s s = 0.02785\ kJ/kg^{-1}$. The pure water mass balance is estimated by

$$\delta = 0.48\ /\ (0.55 - 0.48)\ =\ 6.81.$$

Therefore, $\hat{Q}_A$=0.00275 × 2606.65 + 6.81 × 0.00275 − (1 + 6.81) 0.00275 × 168=5.0071kW.

The heat rejection rate out of the condenser is $\hat{Q}_C$ =0.00275×(2450.80−176) = 5.2687kW. The heat rate added to the evaporator is the cooling consequence created by the absorption cooling system of $\hat{Q}_E$ =0.00275×(2606.65−115.30) = 5kW. The expansion valve mass balance is

$$\tilde{m}_9 = \tilde{m}_8 = \tilde{m}_r = 0.00275 kJ/kg^{-1}.$$

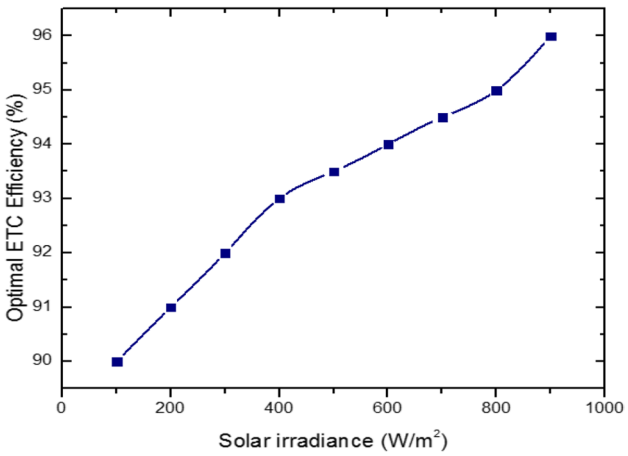


Fig. 7 The finest ETC efficiency validation using GARIC-HACABO method

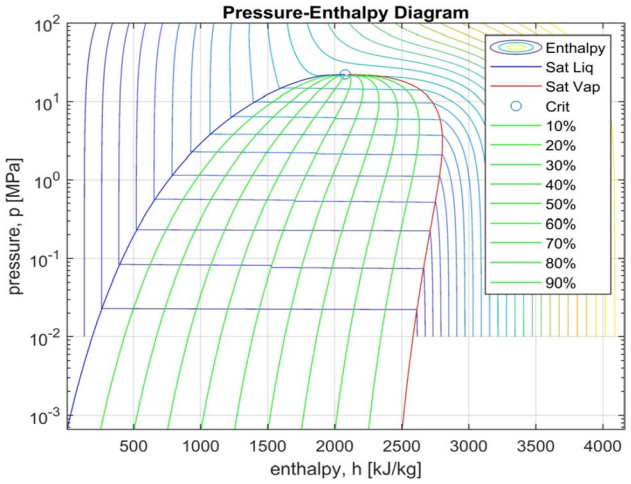


Fig. 8 The working process of pressure-enthalpy diagram for LiBr-H_2O

The cooling effect created by the absorption cooling system of evaporator is

$$\hat{Q}_E = 0.00275 \times 2450.80 + 0.02785 \times 117 - 0.0274 \times 175.46 = 5.993\text{kW}$$

The COP of the absorption system is$\text{COP} = 0.82\%$. Moreover, the efficiency of the second law approach is η_s=0.7872%. The single-effect absorption system covers the concentration of LiBr solution at 50–55% and the temperatures of 60–120°C under operating conditions. The small limitation is it has improve the COP up to 0.82%, which is better than other models because of experimental data. Incorporating the intelligent mechanism for monitoring process to monitor the COP changes continuously will provide better tuning results. The data used for the refrigeration cycle analysis are provided in Table 3.

Table 2 LiBr-H_2O enthalpy–mass flow rate-pressure-temperature-concentration

Point	h (kJ/kg^{-1})	$\tilde{m}$(kJ/kg^{-1})	P (kPa)	T (°C)	%LiBr(C)
1	168	0.0274	6.54	21	53
2	168	0.0274	28.50	21	48
3	175.46	0.0274	28.50	52.25	52
4	117	0.02785	28.50	60	55
5	176	0.02785	28.50	21	52
6	176	0.02785	6.54	21	49
7	2450.80	0.00275	28.50	60	48
8	147	0.00275	28.50	28	48
9	115.30	0.00275	6.54	28	52
10	2606.65	0.00275	6.54	78	51

Table 3 Data used in the refrigeration cycle

Symbol	Description	Values
$\hat{T}_G$	Generator temperature	80°C
$\hat{T}_E$	Evaporator temperature	7°C
$\hat{T}_A$	Absorber temperature	25°C
$\hat{T}_C$	Condenser temperature	30°C
C_W	LiBr Concentration of weak solution (%)	55
C_S	LiBr Concentration of strong solution (%)	53

The absorption system is prepared with an adiabatic absorber in contradiction of the $\hat{T}_G$ and $\hat{T}_A$, when the $\hat{T}_C$ is at 30°C along with $\hat{T}_E$ of at 7°C. Also, the value of COP has attained in the range of 0.1 to 0.8. Nevertheless, in the non-operating area, the COP was considered zero. From this, the temperature of the generator is increased from 60 to 90°C; then, the COP also increases. Nevertheless, when the COP increases the temperature of the absorber also increases.

The observation shows that the temperature of the generator is more sensitive to the temperature of the adiabatic absorber. As a consequence of the thermodynamic design and analysis, the absorber temperature is 25°C when the generator temperature is 80°C. Furthermore, the strong solution of the sub-cooler is left by the LiBr-H_2O solution temperature that is simplified by the equivalent outside cooling fluid and it is utilized in the unit of the condenser. The COP of the overall system is estimated by the product of ETC efficiency and thermal COP of the system absorption cycle. Consequently, the total COP efficiency of the scheme is around 0.7872, which is highly efficient.

The hot water temperature is a significant parameter since the constraints of others are permanent by the obtainable early conditions. For the selected ETC and absorption, the system performance is regulated for hot basis temperature for various standards of cooling water temperature, solar intensity, and temperature of chilled water. The heat transfer of each component is shown in Table 4.

The proposed GARIC-HACABO system is compared with the other existing different solar collector–based cooling systems such as parabolic trough collector (PTC) (Mazloumi et al. 2008), flat plate collector (FPC) (Figaj et al. 2019), and parabolic dish (PD) (Bianchini et al. 2019). Those validations are illustrated in Fig. 9. Hence, this comparison result has proved the efficiency of the proposed method. Here, the experimental data has been taken; by performing the proposed model, the attained COP is 0.82%. In addition, the existing methods also executed in the sample platform and the improvement score has been measured. Compared to existing models, the proposed model has shown the improvement of 1.2% than the existing models.

Table 4 The heat transfer for each component

Symbol	Description	Heat value (kW)
$\hat{Q}_E$	Cooling capacity (output power of evaporator)	5
$\hat{Q}_G$	Input heat to the generator	5.993
$\hat{Q}_A$	Absorber heat is discarded to the environment	5.0071
$\hat{Q}_C$	Condenser heat is discarded to the environment	5.2687
COP	Coefficient of performance	0.82

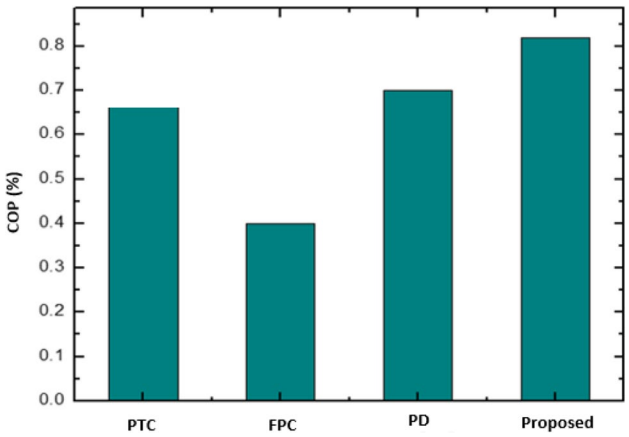


Fig. 9 Comparison of proposed COP with existing solar collector methods

The comparison shows that the proposed method has attained the finest COP value of 0.82 using an efficient ETC, compared with other existing methods. Moreover, the proposed design has been validated with different existing methods such as impedance matching (IM) (Han et al. 2019), metal-organic framework desiccant coated (MFDC) (Xu et al. 2019), solar-linked gradient optimization method (S-GO) (Pang et al. 2019), and ANFIS (Bhowmik et al. 2021), which is illustrated in Fig. 10.

The comparison illustrates that the proposed method has attained superior performance in terms of COP and thermal efficiency. Thus, the overall outcomes show that the proposed framework of the solar-based 5-kW cooling method has attained the finest performance in terms of improved COP value and thermal efficiency rate.

Discussion

The overall analysis of the proposed 5-kW solar-linked vapor absorption system is validated and the state-of-the-art methods are discussed in Table 5. Here, the advantages and limitations along with the methods which were used by

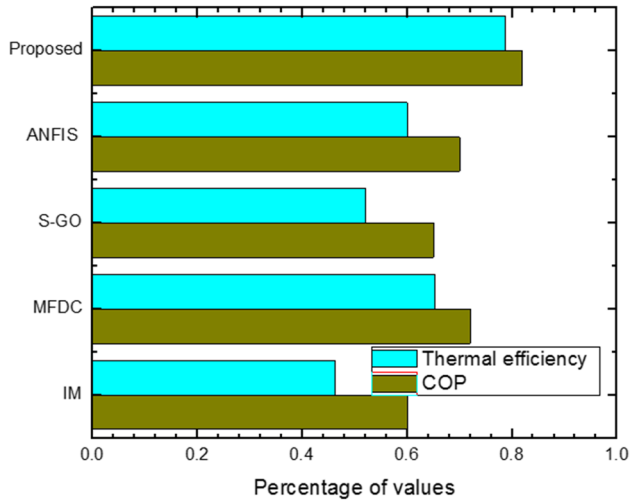


Fig. 10 Comparison of proposed COP with conventional methods

various researchers are compared with the consequences of the proposed technique. The conventional methods have no significant ETC efficiencies and COP values. Moreover, the existing methods have more limitations such as complexity, computational burden, and very less COP rate.

Thus, the discussion illustrates that the proposed GARIC-HACABO in solar ETC for 5-kW vapor absorption refrigeration system has attained the finest outcomes in terms of COP with reduced cost. Moreover, the energy efficiency of the ETC is high; this is because of the hybrid optimization function. In the optimization fitness module, the maximum desired efficiency was imported during the vapor absorption process; the iteration can be processed continuous until it reaches the desired ETC efficiency. Moreover, the regression value in predicting the ETC energy efficiency is 95%; it is also known as prediction accuracy.

Conclusion

In this study, the solar-powered 5-kW LiBr with a water-based vapor absorption system is designed. A petite capacity of 1.4ton intermittent-based vapor absorption refrigeration structure has been achieved utilizing LiBr-H_2O solution depending on the thermodynamic correlation estimations. Moreover, the efficiency of the ETC was evaluated by the proposed GARIC architecture, and the attained efficiency was optimized using the proposed HACABO method. Therefore, the ETC is able to extract high COP from the system. Consequently, the absorption systems are more suitable for refrigeration systems than other old methods. The maximum COP attained from the cycle is in the range of 0.82%. Compared to other associated strategies, the presented model has maximized the COP by 1.2%. Hence, the attained COP rate from the system shows the significant operation of the method under system temperature. Also, the measured thermal efficiency is 0.8%, compared to other approaches; it has maximized the thermal efficiency by 4%. The output power of the evaporator is referred to as the cooling capacity of the system, which has attained 5kW; the heat discarded to the environment by the condenser and absorber is 5.2687 kW and 5.0071 kW, respectively. Also, the main significance of the designed model is high thermal efficiency and COP. Furthermore, this research has provided an exact theoretical model and significant thermodynamic analysis of enhanced vapor absorption–based refrigeration system; also, it has attained less rating of thermal energy basis. Furthermore, the heat input to the system is applied through a hybrid solar collector. This solar-based connection is utilized not only for heating purposes, but also for cooling functions. Moreover, this cooling method will assist the function engineers in attaining specific and rapid forecasting of the system performance. Moreover, the proposed design is validated with the

Table 5 State-of-the-art methods over proposed technique

Author	Techniques	Advantages	Limitation
Han et al. (2019)	Impedance matching	Efficiency is improved	The COP value is very less
Xu et al. (2019)	Metal-organic framework desiccant coated	Rated cooling is possible by this approach; high priority has been attained	Complexity is high and computational burden
Nourafkan et al. (2019)	Binary nanofluid	Stability is achieved	Very less COP while compared with the other techniques
Ye et al. (2019)	Sub-coolant method	Flow rate and temperature value is enhanced	High cost
Pang et al. (2019)	Solar-linked gradient optimization method	Pollution is reduced	Appropriate outcomes are not possible
Bhowmik et al. (2021)	ANFIS	Error rate is less	Effective design cannot achieved
Proposed	GARIC- HACABO	COP is high, reduced cost, improved design efficiency	--

experiment data, so average COP has been gained in some cases. In the future, designing the hybrid optimization with hybrid deep learning features will provide a better improvement rate in COP and thermal efficiency.

Author contribution HK, AKB, and RSM have contributed equally to the work.

Data availability Not applicable.

Declarations

Ethics approval All applicable institutional and/or national guidelines for the care and use of animals were followed.

Consent to participate For this type of study, formal consent is not required.

Consent for publication Not applicable.

Competing interests The authors declare no competing interests.

References

Adham AM, Sulaiman MA (2017) Design of a domestic diffusion absorption refrigeration system using evolutionary algorithm. Int Rev Mech Eng (IREME) 11(11):803

Akbarzadeh A, Wadowski T (1996) Heat pipe-based cooling systems for photovoltaic cells under concentrated solar radiation. Appl Therm Eng 16(1):81–87

Arshad MU, Ghani MU, Ullah A, Güngör A, Zaman M (2019) Thermodynamic analysis and optimization of double effect absorption refrigeration system using genetic algorithm. Energy Convers Manag 192:292–307. https://doi.org/10.1016/j.enconman.2019.03.083

Asadi J, Amani P, Amani M, Kasaeian A, Bahiraei M (2018) Thermo-economic analysis and multi-objective optimization of absorption cooling system driven by various solar collectors. Energy Convers Manag 173:715–727. https://doi.org/10.1016/j.enconman.2018.08.013

Assilzadeh F, Kalogirou SA, Ali Y, Sopian K (2005) Simulation and optimization of a LiBr solar absorption cooling system with evacuated tube collectors. Renew Energy 30(8):1143–1159

Bhowmik M, Muthukumar P, Anandalakshmi R (2021) Experimental study of coupled heat and mass transfer phenomena between air and desiccant in a solar assisted thermal liquid desiccant system. Int J Therm Sci 162:106795

Bianchini A, Guzzini A, Pellegrini M, Saccani C (2019) Performance assessment of a solar parabolic dish for domestic use based on experimental measurements. Renew Energy 133:382–392

Chidambaram LA, Ramana AS, Kamaraj G, Velraj R (2011) Review of solar cooling methods and thermal storage options. Renew Sust Energ Rev 15(6):3220–3228

Ebrahimnia-Bajestan E, Moghadam MC, Niazmand H et al (2016) Experimental and numerical investigation of nanofluids heat transfer characteristics for application in solar heat exchangers. Int J Heat Mass Transf 92:1041–1052. https://doi.org/10.1016/j.ijheatmasstransfer.2015.08.107

Fekadu G, Subudhi S (2018) Renewable energy for liquid desiccants air conditioning system: a review. Renew Sust Energ Rev 93:364–379

Ferreira CI, Kim DS (2014) Techno-economic review of solar cooling technologies based on location-specific data. Int J Refrig 39:23–37

Figaj R, Szubel M, Przenzak E, Filipowicz M (2019) Feasibility of a small-scale hybrid dish/flat-plate solar collector system as a heat source for an absorption cooling unit. Appl Therm Eng 163:114399. https://doi.org/10.1016/j.applthermaleng.2019.114399

Fong KF, Chow TT, Lee CK, Lin Z, Chan LS (2010) Comparative study of different solar cooling systems for buildings in subtropical city. Sol Energy 84(2):227–244

Habibollahzade A, Mehrabadi ZK, Markides CN (2021) Comparative thermoeconomic analyses and multi-objective particle swarm optimization of geothermal combined cooling and power systems. Energy Convers Manag 234:113921. https://doi.org/10.1016/j.enconman.2021.113921

Han Y, Li M, Wang Y, Li G, Ma X, Wang R, Wang L (2019) Impedance matching control strategy for a solar cooling system directly driven by distributed photovoltaics. Energy 168:953–965

Julien K, Spano WF, Kibele R (2016) Agent for the formation of channels in an entrained polymer, entrained polymer containing such an agent, process for producing such an entrained polymer and product containing the same. U.S. Patent Application No. 15/897,897

Kadyan H, Berwal AK (2018) Experimental investigation and analysis of solar refrigeration system. Int J Appl Eng Res 13(13):11349–11353

Kadyan H, Berwal AK, Mishra RS (2018) A new approach to analyze the performance of flat plate collector. J Alt Energ Sour Technol 9(2):39–49

Kalkan N, Young EA, Celiktas A (2012) Solar thermal air conditioning technology reducing the footprint of solar thermal air conditioning. Renew Sust Energ Rev 16(8):6352–6383

Ker YB, Chyau CC, Wang HE, Chen KC, Peng RY (2012) Feasibility of pressure swing adsorption (PSA) plus thermal swing adsorption (TSA) technology for odorant separation-a process design approach. Int J Chem React Eng 10(1):A75. https://doi.org/10.1515/1542-6580.2665

Khosravi A, Syri S (2020) Modeling of geothermal power system equipped with absorption refrigeration and solar energy using multilayer perceptron neural network optimized with imperialist competitive algorithm. J Clean Prod 276:124216

Kim DS, Infante Ferreira CA (2008) Solar refrigeration options–a state-of-the-art review. Int J Refrig 31(1):3–15

Mazloumi M, Naghashzadegan M, Javaherdeh K (2008) Simulation of solar lithium bromide–water absorption cooling system with parabolic trough collector. Energy Convers Manag 49(10):2820–2832

Nourafkan E, Asachi M, Jin H, Wen D, Ahmed W (2019) Stability and photo-thermal conversion performance of binary nanofluids for solar absorption refrigeration systems. Renew Energy 140:264–273

Pang W, Yu H, Zhang Y, Yan H (2019) Solar photovoltaic based air cooling system for vehicles. Renew Energy 130:25–31

Patel CD, Bash CE, Beitelmal AH (2009) Smart cooling of data centers. ASME 2003 International electronic packaging technical conference and exhibition. American Society of Mechanical Engineers Digital Collection

Salahuddin M, Alam K, Ozturk I, Sohag K (2018) The effects of electricity consumption, economic growth, financial development and foreign direct investment on CO2 emissions in Kuwait. Renew Sust Energ Rev 81:2002–2010. https://doi.org/10.1016/j.rser.2017.06.009

Tabor DP, Roch LM, Saikin SK et al (2018) Accelerating the discovery of materials for clean energy in the era of smart automation. Nat Rev Mater 3:5–20. https://doi.org/10.1038/s41578-018-0005-z

Ventas R, Lecuona A, Zacarías A, Venegas M (2010) Ammonia-lithium nitrate absorption chiller with an integrated low-pressure compression booster cycle for low driving temperatures. Appl Therm Eng 30(11-12):1351–1359

Xu F, Bian ZF, Ge TS, Dai YJ, Wang CH, Kawi S (2019) Analysis on solar energy powered cooling system based on desiccant coated heat exchanger using metal-organic framework. Energy 177:211–221

Ye X, Liu L, Li Z (2019) Performance analysis of solar absorption-subcooled compression hybrid refrigeration system in subtropical city. Front Energy 13(1):185–192

Zambolin E, Col DD (2010) Experimental analysis of thermal performance of flat plate and evacuated tube solar collectors in stationary standard and daily conditions. Sol Energy 84(8):1382–1396

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360442883

A REVIEW OF THE DRIP IRRIGATION SYSTEM

Article · July 2022

DOI: 10.55953/JERA.2022.1103

CITATIONS
0

READS
41

5 authors, including:



Ritu Raj Ahirwar
Delhi Technological University
31 PUBLICATIONS **78** CITATIONS

SEE PROFILE



Rahul Kumar Meena
Delhi Technological University
10 PUBLICATIONS **6** CITATIONS

SEE PROFILE

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

Wind effects on tall buildings View project

All content following this page was uploaded by Rahul Kumar Meena on 09 May 2022.

The user has requested enhancement of the downloaded file.

http://dx.doi.org/10.55953/JERA.1103

A REVIEW OF THE DRIP IRRIGATION SYSTEM

Bhagya Jayant[1], Kunal Dahiya[1], Ark Rukhiyar[1] , Ritu Raj[2] & Rahul Kumar Meena[3*]

[1] Undergraduate Student, Department of Civil Engineering, Delhi Technological University, Delhi, India

[1] Post Graduate Student, Department of Civil Engineering, Delhi Technological University, Delhi, India

[2] Assistant Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India

[*] Corresponding author rahul.08dtu@gmail.com

Abstract- Irrigation can be defined as the process of applying artificial water to soil or land in order to promote the growth of agricultural crops such as corn and wheat. Water irrigation has a number of advantages, including an increase in crop yield, protection against famine, revenue generation, and the avoidance of mixed cropping. It also has advantages in navigation, hydroelectric power generation, and the creation of employment opportunities. It should be noted that irrigation has both advantages and disadvantages, just like everything else. Irrigation has a number of disadvantages, including the waste of irrigation water, the formation of marshy lands, damp weather, and the loss of valuable agricultural land. Drip irrigation is one of the surface and subsurface irrigation methods that we will discuss in this paper. We will also discuss the various effects and advancements in this technique that have occurred as a result of the research conducted. Last but not least, the significance and necessity of this system will be discussed.

Key words:- Drip, Irrigation, Artificial Irrigation, Efficiency, Framers

Introduction

Drip irrigation system is a type of micro-irrigation system that is widely used around the world to improve crop yields and to increase crop yield potential. It is the most recent irrigation method to be developed. It is used as a surface and subsurface irrigation method, depending on the situation.

In surface irrigation, water is distributed to the agricultural land through small channels that flood the area to a depth determined by the amount of water required. Hydraulic applications and

http://dx.doi.org/10.55953/JERA.1103

distribution can be accomplished either by gravity or by pumping. When used on soils with low to moderate infiltration capacities and on lands with uniform terrain, this method is effective. It is further divided into three categories: flow and lift irrigation; perennial irrigation; and flood irrigation. Farmers have traditionally used this method to harvest their crops. However, because of the significant disadvantage of this practice was found to be ineffective due to water loss due to evaporation in the middle of the process.

It is also possible to use drip irrigation in conjunction with subsurface irrigation, which is another option. Subsurface irrigation is a type of irrigation in which water flows underground and nourishes plant roots through capillary action. Water is delivered directly to the roots, with no excess water flowing anywhere else in the system. The network is made up of three types of pipes: main pipe, sub main pipe, and lateral perforated pipe. This method is best suited for soils with high water permeability. They can be further divided into two categories: natural sub irrigations and artificial sub irrigations. Artificial irrigation, on the other hand, is more expensive and is therefore only recommended in areas with high returns.

When people first started using this technique, they encountered some difficulties determining when to allow the water to flow. Gradually, they discovered the best method for determining the best time of day to allow water to flow and the length of time it should flow for the best crop yield results.

When people first started using this technique, they encountered some difficulties determining when to allow the water to flow. Gradually, they discovered the best method for determining the best time of day to allow water to flow and the length of time it should flow for the best crop yield results.

LITERATURE REVIEW

There is a plethora of literature on agricultural technology adoption. Drip irrigation adoption, on the other hand, is contingent on a variety of demographic, regional, and other factors, such as government incentives. As a result, according to the study's goal, the suggested study integrates certain relevant literature. Precision farming, according to Yahasaswini Sharma et al (2015), will boost productivity, lower production costs, and reduce the impact of farming on the environment. Precision farming, according to Crookston (2006), is one of the top ten agricultural revolutions in history, and it entails

http://dx.doi.org/10.55953/JERA.1103

using the correct management strategies in the right area, at the right time, and at the right place. According to Mulla et al (1996), precision farming has several advantages, including increased efficiency in farm management inputs, increased agricultural output, improved crop quality, and less fertilizer transportation. Agriculture was linked to socioeconomic trends in underdeveloped nations by Pinaki Mondal et al (2008). The scope, current state, and tactics for precision agricultural adoption in India and other developing nations are also discussed in this research. Khondoker A. Mottaleb (2018) explored the perception and acceptance of new agricultural technology in developing nations. The said study highlights the significance of poverty among farmers in developing countries' rural regions in making technology adoption decisions. According to a study by Neşe Üzen et al on the importance of micro-irrigation in modern agriculture, the micro irrigation system is successful for horticulture in any weather condition. This will result in higher crop yields, better fertilizer use, less salinization, and no crop disease, as well as lower labour costs. P. Suryavanshi et al. (2015) recommended micro-

irrigation as sustainable agriculture since it saves water and increases crop output while lowering cultivation costs. When compared to surface irrigation, all crops farmed with drip irrigation have higher production. Drip irrigation's prospects, possibilities, and limitations in India were explored by Khusro Moin et al (2018); they also discussed the economic state of water in India, as well as a comparison between surface irrigation and drip irrigation in terms of cost-benefit analysis, agricultural yield, labor, and efficiency. It also examines the rate of drip irrigation adoption in India, concluding that the reason for farmers' reluctance to embrace drip irrigation is the high initial cost, which is prohibitive for small and medium farms. According to Namara et al. (2001), the adoption of micro-irrigation technologies involves three essential factors: (a) the technology must be economically and technically feasible; (b) farmers must be aware of the technology, and (c) farmers must be able to access it. With the aforementioned literature research, we can see the limits and reasons why farmers in different parts of India are not using drip irrigation.

METHODOLOGY

Drip irrigation is a technique in which water and fertilizer are applied slowly and directly to the root zone of the plant in order to reduce evaporation and seepage losses. This is accomplished through

http://dx.doi.org/10.55953/JERA.1103

the use of specially designed emitters and drippers. This method is best suited for row crops and orchards, such as tomatoes, grapes, corn, cauliflower, cabbage, and other similar crops, as well as orchards.

Using drip irrigation has a number of advantages, including the fact that it requires little water, has low evaporation losses, produces the fastest rate of vegetative growth, wets the soil surface the least amount of time, and thus reduces the occurrence of diseases caused by damp conditions. There is no need to level the land. There is no soil erosion taking place. There is less labour required. One example is Israel, which was experiencing severe water shortages prior to the implementation of drip irrigation. After implementing drip irrigation, Israel's water supply was significantly improved.

And the disadvantages include a lot many too. It is little expensive, around 30 – 40k per acre or may vary through states/ locality with subsidies imposed on them. Requires extra training apart from notion of general farming. Government is doing its best to raise awareness through camps or books on this newly adopted system. Have to redo, good qualities pipes have general life span of 5 years, which may be exceeded maximum to 8 years. After which, the entire pipe system needs to be removed and changed with the new pipes. This is tedious and time taking. Does not

offer frost protection. Needs regular flushing and supervision. High skill is required in the design, installation, operation and maintenance.

Drip irrigation system is not traditionally irrigation, therefore, farmers need to learn the technique which takes little training (now also provided by government camps and education). Water is saved by minimizing evaporation. Low soil erosion occurs. The main benefit of drip irrigation system is that water is directly at the root zone. Surface drip or surface. Cheap pipes made in possible. Subsidies can be given by the government for installation of drip system. Proper guidelines are given for its entire set – up this is for outlet system.

It saves up to 20 to 50% of water, therefore saving fertilizers like nitrogen gas which are directly sent to the pipes, saving up electricity charges and also recharges the rainwater. Nowadays, new techniques used to resist the rodent incubation involves setting the pipes' plastic layer with posion, so that when any rodent bites it, gets poisoned and dies on spot

http://dx.doi.org/10.55953/JERA.1103

CONCLUSIONS

From various research papers reviewed above as well other secondary research it may be concluded that:

- Farmers with less income or small farmers who are marginalized have less tendency to obtain drip irrigation systems in any form. The major reason for it is the high initial costs and short lifespan of pipes in this system. Also, since most of these small farmers do not have large enough land holdings, therefore investing in drip irrigation techniques becomes not viable for them.
- It may also be concluded that big farmers having large landholdings are more willing to acquire this technique as it saves them to labor, water, electricity, and fertilizer costs. It optimizes available water for them as evaporation loss is less and water is dripping directly at the root of the plant.
- Although, for a country like India, drip irrigation is very beneficial and must be adopted. However, when the matter of the world is concerned, it has its own complications. Some countries like New Mexico which face deficit irrigation develop stresses in plants thereby making the crop more prone to diseases and attacks causing environmental harm. For such areas, this technique takes a dip as the crop yields are reduced as water never reaches the water deficit areas. For many such places as such, water-saving techniques are used which may reduce the downstream flow of water and raise production.
- Drip irrigation was found to be better than most of the other irrigation methods due to its multiple advantages and adaptability in the current scenario in the irrigation system. It is simple and needs little training to implement. The government in many states also promoted the procurement of the same by providing subsidies.

Data Availability Statement

All data, models and code generated or used during the study appear in the submitted article.

Funding Statement

http://dx.doi.org/10.55953/JERA.1103

Authors would like to express their sincere gratitude to Delhi Technological University, Delhi, India for providing research facilities to conduct this research work.

Acknowledgments:

The authors acknowledge the Department of Civil Engineering and Delhi Technological University, Delhi

Conflict of Interest:

The authors declare that there are no conflicts of interest.

REFERENCES

[1] Tiwari, K. & Mal, P & Singh, R & Chattopadhyay, Arup. (2019). Drip irrigationpaper in Agric. water management. Agricultural Water Management. 38. 91-102.

[2] Asiegha, J.E., 1991. Responses of tomato and egg plant to mulching and nitrogenfertilization under tropical conditions. Sci. Horti. 46, 33±41.

[3] Call, R.E., Courter J.W., 1989. Response of bell pepper to raised beds, black plastic mulch, spunbonded row cover and trickle irrigation. Proceedings of 21st NAPC, pp. 140±146.

[4] Decoteau, D.R., Kasperbauer, M.J., Hunt, P.G., 1989. Mulch surface color effects yield of fresh-market tomatoes. J. Am. Soc. Hort. Sci. 114, 216±220.

[5] Doorenbos, J., Pruitt, W.O., 1977. Guidelines for predicting crop water requirements. Irrigation and Drainage Paper No. 24 (revised), FAO, Rome

[6] Sharma, Y., & Ashoka, P. PRECISION FARMING AND USE OF SENSORS IN HORTICULTURE. 2. Crookston, R. K. (2006). A top 10 list of developments and issues impacting crop management and ecology during the past 50 years. Crop science, 46(5), 2253-2262. 3.

[7] Mulla, D. J., Perillo, C. A., & Cogger, C. G. (1996). A site-specific farm-scale GIS approach for reducing groundwater contamination by pesticides. Journal of Environmental Quality, 25(3), 419-425.

http://dx.doi.org/10.55953/JERA.1103

[8] Larson, N., Sekhri, S., & Sidhu, R. (2016). Adoption of water-saving technologyin agriculture: The case of laser levelers. Water Resources and Economics, 14, 44- 64.

[9] PANDYA, P. A., & DWIVEDI, D. K. Constraints in Adoption of Drip Irrigation.

[10]Namara, R., Upadhyay, B., & Nagar, R. K. (2005). Adoption and impacts of microirrigation technologies: Empirical results from selected localities of Maharashtra and Gujarat States of India (Vol. 93). IWMI. 20.

Received: 18 March 2022
Revised: 11 April 2022
Accepted: 26 April 2022
Journal of Engineering Research and Application

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

Article

A Structural Approach on Enablers of IoT for Sustainable Development of Smart Cities

Ayusha Thakur[1], Satyabrata Aich[2] and Rohit Kumar[3]

[1] Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India
[2] Wellmatix Co., Ltd., Changwon 51395, Korea
[3] Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India

Abstract: Smart cities will undoubtedly be the distinguishing characteristic of human geography in the twenty-first century. Throughout 1.3 million people move to cities every week around the world, and it is estimated that by 2040, cities will house 65 percent of the world's population. While the world's largest cities now contribute for 60% of global GDP, this percentage will continue to climb as cities get larger and smarter. According to experts, cities will account for up to 80% of future economic development in developing regions. Smart cities are no longer the wave of the future; they have here and are rapidly expanding as the Internet of Things (IoT) grows. With dozens of towns throughout the world, the smart city sector is expected to grow into a massive business as time goes on. Cities have been an increasingly crucial engine of the global economy and wealth over time. As a result, it is critical to guarantee that they are optimized to maximize efficiency and sustainability while also ensuring that each citizen's quality of life is improved. We can describe the need for smart cities through this project, and it also shows us how IoT technology and smart city enablers may be deployed in urban settings to help cities perform better for their residents and achieve overall sustainable growth. This initiative will educate urban planners, researchers, ICT professionals, and other city officials about the facilitators of IoT for smart cities and their long-term growth.

Keywords: Enablers of IoT; Interpretive Structural Modeling; Smart Cities; Sustainable Development

1. Introduction

Among numerous definitions of a smart city, in general terms, we can define a smart city as a city that utilizes IoT sensors, actuators and technology to connect components across the city that impacts every layer of a city which is from underneath the streets to the air that citizens breathe.Data from all segments is evaluated, and patterns are then created from the information gathered. A smart city is a framework made up mostly of information and communication technologies (ICT) that is used to position, advance, and promote sustainable development methods in order to track and address urbanization's progress and issues.

1.1. IoT in Smart City

Since the concept of a smart city was first established, Internet of Things (IoT) technology has been a major pillar of smart city growth. As technological advancements evolve and more countries adopt next-generation connectivity, it will continue to increase and have a greater impact on how we live.

"By 2025, there will be more than 75.44 billion linked IoT devices," according to the Improving Internet of Things (IoT) Security with Software-Defined Network (SDN) research. The Internet of Things is predicted to evolve into one of the brightest collective and collaborative systems in history, with over 7.33 billion mobile users by 2023 and over

© 2022 by the author(s). Distributed under a Creative Commons CC BY license.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

1,105 million connected wearable device users by 2022. Cities must comprehend the benefits and prospects of the Internet of Things for Smart Cities since there is so much promise and opportunity across a wide variety of sectors, including urban transportation, security, sustainability, maintenance, healthcare, and management."

In a smart city, the Internet of Things is one of the main components that binds everything together. From sophisticated interconnectedness to becoming one of the key building elements of next-generation, smart city development, residents and governments will be connected in unprecedented ways.

1.2. IoT for Sustainable Development of Smart City

Urbanization is a continual phenomenon, and cities now house more than half of the world's population. By 2050, this percentage is predicted to be much higher. Over the next three decades, the overall population growth and growing urbanization will add many billions more people to cities. To keep up with the fast development that is stressing our cities' resources, environmental, social, and economic sustainability has become a requirement. Therefore, according to the UN, "in September 2015 at the United Nations, 189 nations agreed on the Sustainable Development Goals (SDGs) agenda."

Smart cities using IoT are using analytics and data from sensors to design better civic amenities and deliver better standards of living and cost-effective public services. With the rise in population in cities and the corresponding increased stressors on the natural and civic resources, civic administrators are coming up with interesting ways to utilize their data, IT infrastructure and using IoT applications in smart cities to deploy advanced models that will help them solve the challenges of urban planning efficiently and effectively.28 percent of people worldwide will be concentrated in cities with at least 1 million inhabitants due to which Smart cities leveraging technological advances to deliver a superior quality of living to the residents have become a critical public policy planning element. With the rise in the global population, it has become imperative for governments to design a sustainable and environmentally friendly future, to ensure the well-being of the people and curtail global warming. Major global cities have pledged to adopt sustainable practices and ensure significant cut-downs along critical dimensions and support the UN Sustainable Development Goals progress.

Following a comprehensive analysis of numerous literatures relevant to our research, we discovered that there are few attempts to link the enablers of IoT for smart cities and sustainability together. Furthermore, numerous frameworks and roadmaps have been offered to attain their desired goals, but no study has utilised ISM methodology to determine the driving and dependent capabilities of the selected enablers. With this research gap in mind, the goal of this study was to determine the structural link between IoT enablers for Smart Cities for Sustainable Development.

Following an exhaustive literature research and expert consultation, we have come up with 11 IoT enablers for Smart Cities and their link with sustainable development. The study used Interpretive Structural Modeling (ISM) to investigate the relationships between the extracted factors and their interdependencies, leading to the creation of a driver-dependence diagram to access the factors' driving and dependence power.

The research is divided into five components. To begin, Section 2 gives an overview of the link between IoT and its enablers based on a review of the literature and expert perspectives. Section 3 discusses the suggested ISM approach. The results of the ISM technique are interpreted in Section 4, and the management implications are discussed in Section 5.

2. Literature Review

Simon Elias Bibri explains how the Internet of Things (IoT) is one of the key components of the ICT infrastructure of smart sustainable cities as an emerging urban development approach due to its great potential to advance environmental sustainability in his

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

paper "The IoT for Smart Sustainable Cities of the Future: An Analytical Framework for Sensor-Based Big Data Applications for Environmental Sustainability." His paper aims to review and summaries relevant literature in order to identify and discuss futuristic sensor-based big data applications enabled by the Internet of Things for environmental sustainability, as well as related data processing platforms and computing models, in the context of smart sustainable cities of the future. His paper also identifies the major challenges associated with the Internet of Things and big data analytics. Furthermore, it has also looked into the possibility of using big data applications to supplement the informational landscape of smart sustainable cities in order to reach the requisite level of environmental sustainability.

The proposed framework was tested in empirical research to add depth to studies in the field of smart sustainable cities, allowing the paper to inform urban planners, scholars, ICT experts, and other city stakeholders about the environmental benefits that can be gained from implementing smart sustainable city initiatives and projects based on IoT-based smart and sustainable city applications. An intelligent network of connected items and equipment that continuously transfer data utilizing wireless technology and the cloud is an important aspect of the ICT framework. Individuals, society, and industries can use cloud based IoT apps to receive, analyse, and manage data in real time to help them make better decisions that improve their quality of life. We interact with smart city ecosystems in a variety of ways, including through smartphones and mobile devices, connected cars and homes, and by combining these devices and data with a city's physical infrastructure and services, we can save money and increase sustainability. As a result, smart city solutions are divided into the following categories in order to transform the urban landscape:

1. Infrastructure

All of the core elements associated with a smart city, including smart people, smart transportation, smart economics, smart living, smart government, and smart environment, are built on top of smart infrastructure. A smart infrastructure is a cyber-physical system that allows for the integrated management of all of its components using various technology tools to assemble and analyse data in order to achieve efficiency, sustainability, productivity, and safety goals. It is a system that monitors, measures, analyses, communicates, and acts on data acquired by sensors and uses a data feedback loop to improve decision-making. This infrastructure receives data, interprets information, and takes necessary action totally autonomously, that is, without the need for human involvement, and adjusts to changing situations. Smart networks, smart buildings, smart public infrastructure, and smart beaches are typically included in this category. Smart buildings are those that have been constructed with modern installation and technological systems, allowing for the automation of various internal activities such as heating, ventilation, lighting, security, and other building systems. Most of these components have one thing in common: they are connected, and they provide data that may be used intelligently to ensure the best use of resources and increase performance.

2. Governance

One of the hallmarks of smart cities is smart governance, which has its roots in e-government, good governance principles, and the concepts of citizen participation and involvement in public decision-making. It's about the future of government services, increased efficiency, community leadership, mobile working, and continuous improvement through innovation and the use of technology to assist and promote better planning and decision-making. It's all about enhancing democratic procedures and altering the delivery of public services, which includes e-government, the efficiency agenda, and mobile working.

3. Utilities

Sensors, improved metres, renewable energy sources, digital controls, and analytic tools are used in smart energy management systems to automate, monitor, and optimise

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

energy distribution and usage. By balancing the interests of the various stakeholders involved, such systems optimise grid functioning and usage (consumers, producers and providers). Distributed renewable generation, microgrids, smart grid technology, energy storage, automated demand response, virtual power plants, and demand-side developments such as electric vehicles and smart appliances are all examples of smart energy infrastructure advances. Such advances enable community-based energy monitoring programmes and improve building energy efficiency by extending a network of intelligent energy devices across a city and providing a thorough view of energy consumption patterns. Smart grids are an important part of smart energy infrastructure. A smart grid is a "electricity delivery system connected with ICT from point of generation to point of consumption for increased grid operations, customer services, and environmental benefits."

4. Mobility

Smart mobility is best defined as methods for reducing traffic congestion and promoting faster, greener, and less expensive transportation solutions. Most smart mobility solutions employ data about mobility patterns gathered from a range of sources to assist enhance traffic conditions holistically. Individual mobility systems, such as bicycle sharing, carpooling, vehicle sharing, and, more recently, on-demand transportation, are examples of smart mobility systems. New disruptive business models based on sharing, convenience, and technological use are emerging. To assess the influence of such models on traffic congestion and car usage in cities, more research is needed. Intelligent transportation systems integrate a city's complete multimodal transportation options, including both individual and group mobility. Intelligent transportation systems efficiently integrate a city's whole multimodal transportation options, including both individual mobility and mass transit. A network of sensors, GPS-tracked public transportation, dynamic traffic lights, passenger information panels, automatic vehicle registration plate readers, closed-circuit television systems, navigation facilities, signalling systems, and, most importantly, the ability to integrate live data from most of these sources are all common features of modern intelligent transportation systems. Safety, network management, transportation congestion, environmental performance, accessibility, convenience, and public perception can all benefit from this.

5. Environment

Smart systems for controlling environmental quality, irrigation, garbage, photovoltaics, lighting, weather stations, and water supply are among the smart solutions for the environment. Its goal is to increase energy efficiency and environmental quality in cities. The rate of waste generation is outpacing the rate of urbanisation. 18 Cities are finding it increasingly difficult to source, separate, and utilise various types of waste that may be returned to the consumer life cycle. Waste management usually entails waste collection, transportation, processing, recycling, and disposal. Smart waste management systems decrease waste at the source, classify waste types, and establish procedures for proper trash disposal. Such devices could be utilised to turn garbage into a resource and generate electricity. Closed-loop economies can be created by converting waste into a resource using such technologies. Their main advantages are increased garbage collection, pickup, separation, reuse, and recycling efficiency.

6. Life

Citizens' smart solutions entail optimising the services provided to them, with the understanding that cohabitants are essential in co-constructing a better city for all. Informational MUPIS, citizen cards, citizen Apps, and social Wi-Fi are examples of these solutions. Its goal is to improve communication between the various parties. Regarding the sustainability of urban areas and their supporting ecosystems, the health and well-being of urban dwellers is of special relevance. Smart cities can improve their ability to employ technologies like big data to make predictions or detect population health hotspots (such as epidemics or health impacts during extreme weather events). Digital health records,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

home health services, and remote diagnosis, treatment, and patient monitoring are all examples of smart health-care management, which translates health-related data into clinical and business insights.

Digital health records, home health services, and remote diagnosis, treatment, and patient monitoring technologies are all examples of smart health-care management, which translates health-related data into clinical and business insights. It also makes it easier to provide health care by utilising intelligent and networked devices that monitor residents' health. With the recent pandemic, smart health approaches such as crowdsourcing to collect data on epidemics, predict epidemic outbreaks, and take necessary precautions, remotely collecting patient health vitals and data for diagnostic purposes, and setting up automated alerts for patients regarding medications and health check-ups are all examples.

Aside from people and traditional urban infrastructure, the following four aspects are critical for smart cities to thrive:

1. Ubiquitous wireless connectivity
2. Open data
3. Steadfast Security
4. Flexible monetization schemes

2.1. Past Literatures concerning IoT Enablers for Smart Cities

In their article, Yonghee Kim, Youngju Park, and Gwangsuk Song[5] employed ISM to determine the enablers influencing IoT service acceptability. Their research concluded that international cooperation and the security of objective data are more important than ever in laying the groundwork for effective IoT services. In addition, efforts should be made to expand and develop IoT services by undertaking research and development and implementing projects through public-private partnerships and the formation of a consultative group. Furthermore, both the government and the business world should concentrate on them.

Pooja Gupta and Vijay Kumar Jain [7] developed a structural model of several enablers that are required to execute G-IoT. From a literature study and expert perspectives, they identified a variety of G-IoT enablers and constructed a structural model of G-IoT enablers using the Interpretive Structural Modeling (ISM) approach.

2.2. Enablers of IoT for Sustainable Development of Smart Cities

1. AIoT

A term used to describe AI in conjunction with the IoT. This is often used when thinking about IoT in an analytics and data collection construct. When adding AI and IoT to make AIoT together, those IoT devices can analyze data used within equipment or devices and make proactive, intelligent, and accurate decisions without the involvement of humans. Eventually, these devices will become smart, communicative, and powerful devices that can process data and make decisions faster and more accurately than ever. Humans are understandably more trusted, and the switch to being more dependent on machine intelligence is not easy. In essence, AIoT, can bridge this gap and make the transition from trusting humans to trusting technology as seamless as possible.AI and IoT also interact with smart office buildings. Companies install sensor technologies that help to manage energy and lighting. Buildings also pick up on the physical movement of people, something essential for security management. Smart buildings can also control access of people through facial recognition. Another way AIoT is being used today is through autonomous vehicles using radars, GPS, and cameras to make real-time decisions based on current driving conditions.

2. Big Data

In a smart city, big data plays a vital role. Cities can discover trends and requirements by analyzing data from IoT devices and sensors. The analysis can assist drivers find a

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

parking place and minimize the amount of road accidents and congestion. Data may also help with crime reduction, smart city lighting, and water and energy systems.

3. Cloud Computing

For analytical reasons, smart cities require a vast volume of data. In their data centers, cloud data systems employ solid-state drives, delete redundant data, and secure data transfer. Payment choices for cloud-based systems are frequently more flexible than those for on-premises data centers. Clouds are essentially low-maintenance and secure for those who utilize them.

4. Predictive Maintenance(PDM)

PDM and wireless power have the potential to make Smart Cities even more intelligent. There are several applications for anything from putting a variety of sensors in metropolitan settings to monitoring traffic patterns, air quality, and the health of key infrastructure. A key challenge of installing strain sensors on a bridge or highway overpass is sourcing the power required for them to operate. Nikola Labs technology can address this challenge by harvesting power from nearby power lines, radio and TV towers, and cell phone communication hubs. When consistent power is provided, long-lived sensors can provide the rich streams of data necessary to generate PdM insights; thereby avoiding catastrophic failures and permitting municipal departments to focus their attention on infrastructure most in need of repair.

5. Preventive Maintenance(PVM)

Similar to predictive maintenance, the goal here is to reduce equipment failures. However, with preventive maintenance, maintenance is performed regularly on the machine to reduce the likelihood of failure rather than doing maintenance when the machine begins to exhibit abnormal behavior . Due to the high-risk nature of manufacturing operations, the only solution for never fail situations has been to over-ensure uptime with redundant equipment and too many parts on hand. While these systems track performance of various components using sensors, there's typically no proactive way to trigger maintenance to avoid downtime.

6. SDGs (Sustainable Development Goals)

The United Nations adopted the 17 Sustainable Development Goals in 2015, which form a comprehensive global framework for ending poverty and protecting the environment by 2030. A recent study found that 84 percent of IoT deployments address the SDGs in some way. The private sector was responsible for 70% of these deployments.

7. Smart Buildings(SB)

Building automation has progressed far beyond HVAC and lighting management systems in recent years. While the notion of smart buildings has been around for a while, with the advent of the Internet of Things and a new generation of intelligent edge devices, the term "smart" has taken on a whole new meaning in the context of building automation. Residential and commercial buildings of all sizes and uses are becoming networked smart ecosystems in which systems monitor and manage far more than just temperature control and lighting. More than merely collecting point data will be done by intelligent edge devices. They'll also collect, analyse, and feed data to edge computing, where powerful predictive analytic engines will allow new levels of management and security while vastly enhancing the overall user experience in the building. These brand-new They'll also collect, analyse, and feed data to edge computing, where powerful predictive analytic engines will allow new levels of management and security while vastly enhancing the overall user experience in the building. These new systems will be more intelligent, self-learning, inventive, and complex.

8. Quality of Service(QoS)

A Quality of Service (QoS) measurement indicates how well a service is delivered. The phrase refers to network performance in the context of IoT connection. In the Internet of Things, quality is a critical aspect in ensuring service quality. The QoS requirements will differ from one application to the next. While framing the application, the quality of service is regarded a vital aspect in the service process.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

The relevance of QoS in IoT has been discussed by Ravi et al. [1.] The authors have summarized the QoS parameters and metrics that must be taken into account at each tier. The authors have organized the available IoT literature into numerous study themes such as standardization, system design and performance, and quality of service, among others.

9. Telematics

A telematics platform is a computer system that handles data transfer across great distances. Many vehicles include consoles that can give GPS and satellite radio services, as an example. In simple terms, it is the use of telecommunications networks to send and receive data from assets located at a distance. Telematics is a vital and lucrative sector of the Internet of Things. Vehicles may be augmented with electronics, connections, and hardware that allow them to talk and interact with other devices when IoT and telematics are combined. Telematics IoT applications are making our roads safer, making driving more fun, and assisting companies with vehicle management. As a result, drivers who use these apps claim a variety of advantages, including theft prevention, decreased insurance costs, and enhanced safety. Telematics IoT applications are making our roads safer, making driving more fun, and assisting companies with vehicle management. As a result, drivers who use these apps claim a variety of advantages, including theft prevention, cheaper insurance costs, and better driving behavior.

10. Range of Applicability

The variety of IoT applications is one of the most fascinating aspects of the technology. While equipment performance should be tailored to a given property type, other IoT devices may be broadly installed at a reasonable cost. Leaks from packaged terminal air conditioners (PTACs), for example, are a regular and costly problem for many real estate firms. Operators are also responsible for maintaining the space temperature of tenant sections. It is unnecessary for real estate businesses to have to hire two different vendors to monitor leaks and space temperature.

11. Network Effect

The network effect is a phenomenon in which the value of a commodity or service increases as the number of persons or participants increases. The network effect may be seen on the Internet. As the internet—one of the most visible examples of the network effect—becomes a more significant part of our lives, it will become increasingly vital for both service providers and users to understand the network effect and its advantages.

3. Methodology

To begin with our research, we identified IoT enablers for smart cities and sustainable development; we analysed previous work and solicited various expert viewpoints and inputs into organised self-interaction matrices (SSIMs). For assessment, we also studied and analyzed related publications and research papers.

Following that, we utilized ISM to create a baseline model of crucial factor connections and MICMAC analysis to categorise the variables. Our ISM looks for connections between items recognised as a concern or a problem through literature study and expert opinion. We used a combination of brainstorming, nominal group tactics, and face-to-face interviews to get expert opinions on how to build a contextual link between crucial components.

We've looked at the IoT enablers for smart cities while keeping sustainable development in mind. A literature study was used to identify variables, and there was no restriction on the number of factors.

Table 1 lists the 11 enablers identified. **Table 2** lists the ISM step followed by the flowchart of the factors as shown in **Figure 1**.

Table 1. Numbers of enablers evaluated in various reports.

S No.	Factors

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

F1	AIoT(AIoT)
F2	Big Data(BD)
F3	Cloud Computing(CC)
F4	Predictive Maintenance(PDM)
F5	Preventive Maintenance(PVM)
F6	Sustainable Development Goals(SDG)
F7	Smart Buildings(SB)
F8	Quality of Service(QS)
F9	Telematics(TM)
F10	Range of Applicability(RA)
F11	Network Effect(NE)

Table 2. ISM steps.

S.No.	Steps	Focus
1	**Establishment of a Structural Self-Interaction Matrix (SSIM)**	"The Driving pairwise relationships among identified critical dimensions of Enablers of IoT and Smart City with Sustainable Development"
2	**Create a reachability matrix**	"To Determine driving and dependent factors"
3	**Level partitioning**	"To Define structural levels (factors level partitioning)"
4	**ISM modeling**	"To Develop an ISM model using a reachability matrix and level partitioning"
5	**MICMAC Analysis**	"To Classify critical dimensions of the enablers of IoT into four factors: drivers, dependents, autonomous factors and linked factors via MICMAC Analysis"

3.1. ISM

ISM is a traditional and extensively used technique utilised by strategic decision makers in numerous businesses. It is used by researchers in various disciplines such as supplier selection, knowledge management, green supply chain management, and energy conservation. When a problem develops, the goal of ISM is to detect and create relationships between aspects influencing decision-making. The problem is then solved by taking

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

into account each factor's driving and dependence capabilities. Experts identify the relationships between components, which finally comprise the research framework.

Table 3 lists the various applications of ISM.

The following steps are involved in modelling: (1) identification of relevant factors based on previous studies and expert opinion; (2) development of an SSIM, then a reachability matrix; (3) creation of a partition level table using the reachability matrix; (4) characterization of relationships between various factors; and (5) identification of uncertainties and subsequent moderations.

Table 3. Applications of ISM.

Techniques	Application
ISM	Interpretive Structural Modeling in the Adoption of IoT Services [**5**]
ISM	Analyzing the factors influencing integrated clean sustainable system [**17**]
ISM	The IoT for Smart Sustainable Cities of the Future: An Analytical Framework for Sensor-Based Big Data Applications for Environmental Sustainability [**1**]
ISM	Factors Affecting ESG towards Impact on Investment: A Structural Approach [**18**]

3.1.1. The SSIM

The SSIMs were completed after extensive investigation and provided us with ISM inputs. The contextual correlations among the 11 criteria that were identified were determined using most opinions on various web platforms. After evaluating the nature, the purpose, and the majority view of the linkages of each difficulty between the variables, we arrived at a conclusion that finalized the contextual relationships.

The contextual interconnectedness between two components I and t) is expressed in one of four ways: (a) if I affects t, "V"; (b) if t influences I "A"; (c) if I and t influence each other, "X"; and (d) if I and t are independent, "O."For example, AIoT, F1(AIoT), has an impact on Predictive Maintenance, F4(PDM), hence the symbol used is V.Because Sustainable Development Goal F6 (SDG) affects Sustainable Buildings Goal F7 (SB), the symbol used is A.The symbols X represent the interaction between AIoT F1 (AIoT) and Big Data F2 (BD).

Table 5 shows the summary of the SSIM. The SSIMs' reachability matrix is discussed in further detail below.

Table 4: Structural Self Interaction Matrix

Factor	**F1**	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**F9**	**F10**	**F11**

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

s											
F1		X	X	V	V	V	V	V	V	V	V
F2			X	V	V	V	V	V	V	V	V
F3				V	V	V	V	V	V	V	V
F4					O	V	V	V	V	V	V
F5						V	V	V	V	V	V
F6							A	A	A	A	A
F7								A	A	A	A
F8									V	X	V
F9										A	X
F10											V
F11											

3.1.2. Reachability Matrix

The four SSIM representations, V, A, X, and O were then replaced by 1 or 0 in the reachability matrix shown below in **Table 5**, as: (a) the symbol "V" in the (i, t) position of the SSIM matrix being substituted by 1 and 0 in the (i,t) and (t, i) positions of the reachability matrix; (b) the symbol "A" in the (i, t) position of the SSIM matrix being substituted by 0 and 1 in the (i, t) and (t, i) positions of the reachability matrix; (c) the symbol "X" in the (i, t) position of the SSIM matrix being substituted by 1 in both the (i, t) and (t, i) positions of the reachability matrix; and (d) the symbol "O" in the (i, t) position in the SSIM matrix being substituted by 0 in both the (i, t) and (t, i) positions of the reachability matrix.

After that, the reachability matrix's transitivity was tested. The term "transitivity" refers to the idea that if factor F1 influences factor F2 and factor F2 influences factor F3, then F1 influences F3. F1's location (I, t) has an effect on F3, and the value becomes 1. For the driving power (DVP), we may compute a factor by summing all numbers in the accommodating row, while V stands for the dependency power (DNP). Following the factoring technique, the final form of the reachability matrix is displayed in **Table 5**. The reachability matrix is also employed in the Level Partitioning process that follows.

Table 5: Reachability Matrix

Factors	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	Driving Power
F1	1	1	1	1	1	1	1	1	1	1	1	11
F2	1	1	1	1	1	1	1	1	1	1	1	11
F3	1	1	1	1	1	1	1	1	1	1	1	11
F4	0	0	0	1	0	1	1	1	1	1	1	7
F5	0	0	0	0	1	1	1	1	1	1	1	7
F6	0	0	0	0	0	1	0	0	0	0	0	1

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

F7	0	0	0	0	0	1	1	0	0	0	0	2
F8	0	0	0	0	0	1	1	1	1	1	1	6
F9	0	0	0	0	0	1	1	0	1	0	1	4
F10	0	0	0	0	0	1	1	1	1	1	1	6
F11	0	0	0	0	0	1	1	0	1	0	1	4
Dependence	3	3	3	4	4	11	10	7	9	7	9	

3.1.3. Level Partition

The reachability and antecedent sets for each element were then constructed using the previously prepared reachability matrix. The elements themselves, as well as factors influenced by other factors, are included in the reachability set. The antecedent set includes both the factors and the factors that influence them. The intersection set consists of components that are shared by both the antecedent and reachability sets. When the antecedent and reachability sets were equal, the technique was repeated, and the top factor was selected. Because the antecedent and reachability sets are equal, F6 occupies level I for Instance. When determining the amount of a factor, six iterations were used. **Table 6** depicts the level divide. There are six levels to each of the 11 criteria. The sixth level is occupied by F1, F2, and F3, while the first level is occupied by F6, and the remaining factors are located between these levels.

Table 6. Level partition.

Factors	Reachability Set	Antecedent Set	Intersection Set	Level
AIoT	1,2,3,4,5,6,7,8,9,10,11	1,2,3	1,2,3	6
BD	1,2,3,4,5,6,7,8,9,10,11	1,2,3	1,2,3	6
CC	1,2,3,4,5,6,7,8,9,10,11	1,2,3	1,2,3	6
PDM	4,6,7,8,9,10,11	1,2,3,4	4	5
PVM	5,6,7,8,9,10,11	1,2,3,5	5	5
SDG	6	1,2,3,4,5,6,7,8,9,10,11	6	1
SB	6,7	1,2,3,4,5,7,8,9,10,11	7	2
QS	6,7,8,9,10,11	1,2,3,4,5,8,10	8,11	4
TM	6,7,9,11	1,2,3,4,5,8,9,10	9,12	3
RA	6,7,8,9,10,11	1,2,3,4,5,8,10	8,11	4
NE	6,7,9,11	1,2,3,4,5,8,9,10,11	9,11	3

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

The digraph and level partition table were used to create the ISM depicted in **Figure 1**. This digraph shows the interrelationships between items at edges and nodes while excluding transitive interactions. The ISM was derived using the digraph's combinative information.

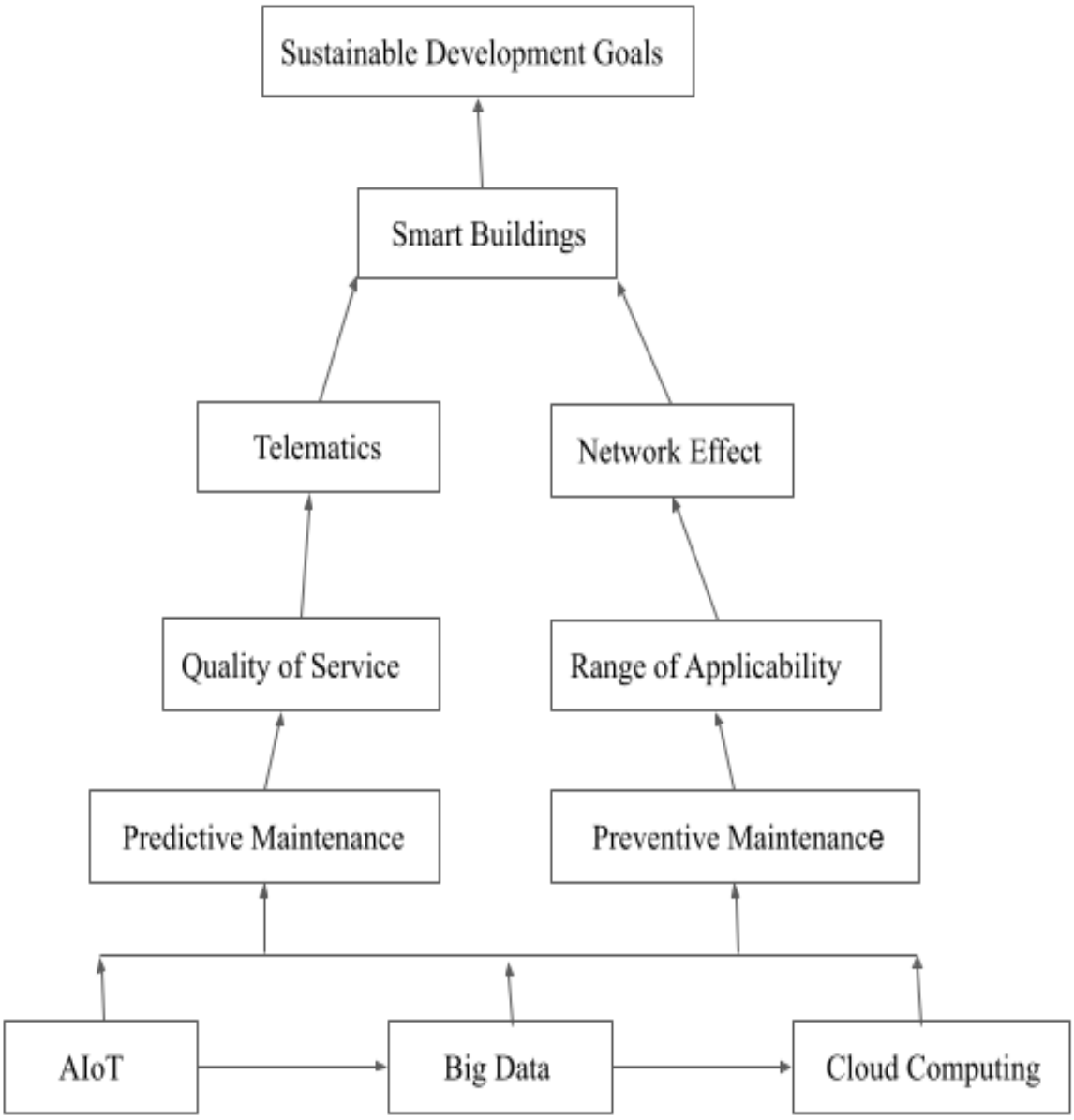


Figure 1 ISM Model

3.2. MICMAC Analysis

We need factor dependence and driving powers as inputs to perform the analysis, which then requires factor dependence and driving powers as inputs to categorise the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

factors into four types (**Figure 2**): Autonomous variables (weakly dependent and driving factors) are depicted in the first quadrant. Linkage variables (factors with strong dependency and driving powers) are displayed in the third quadrant, while dependent variables (factors with high dependence but weak driving capabilities) are displayed in the second quadrant. The third quadrant contains driving variables (factors with low dependency powers but high driving powers).

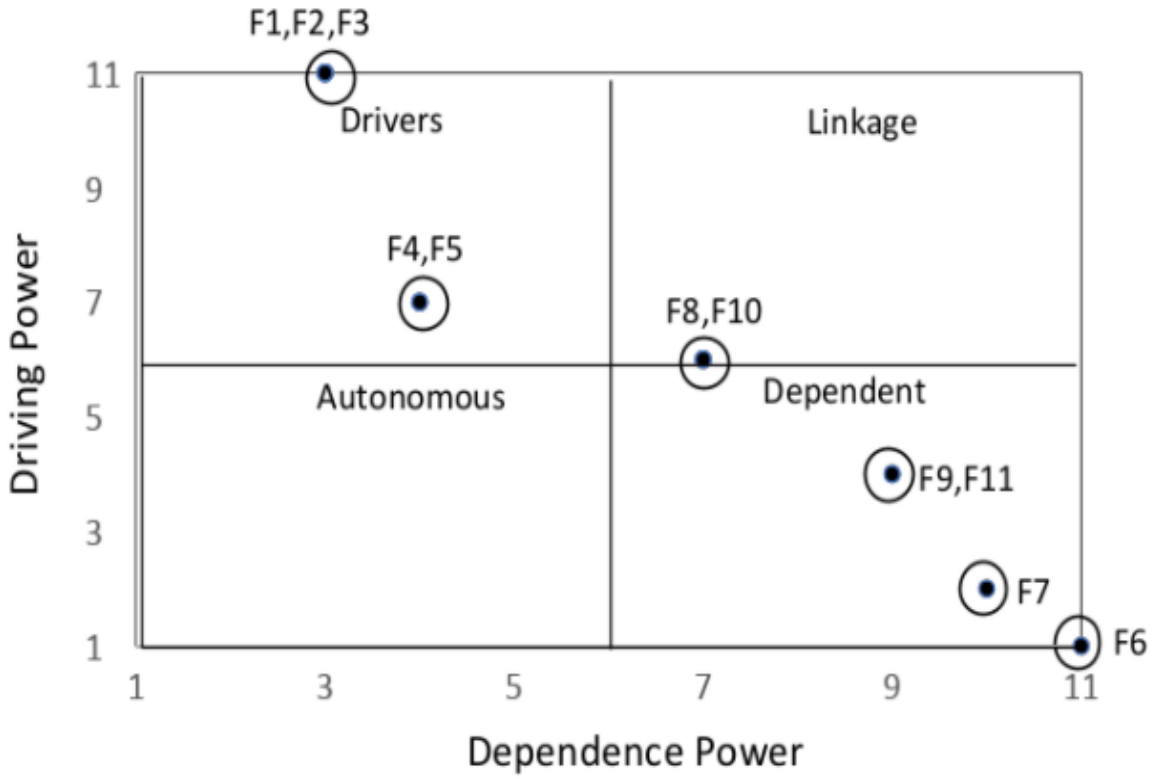


Figure 2. MICMAC analysis

5. Conclusion

Our system has systematically analyzed and constructed the relationships among enablers of IoT for Sustainable Development of Smart Cities. We first identified critical factors affecting the adoption through MICMAC analysis and have systematically analyzed and constructed the relationships among factors relating. We have derived factor dependence and driving powers through the analysis. The ISM split all factors into six levels. AIoT, Big Data and Cloud Computing were observed to be the key enablers as they are put up at the bottom in the hierarchy.

In **Figure 1** The ISM is split into six tiers. The consideration of AIoT (AIoT)(F1), Big Data (BD)(F2) and Cloud Computing (CC)(F3) are the fundamental parts of enablers of IoT (which were positioned at the bottom of the hierarchy) while Sustainable Development Goals (SDG)(F4) being on the top as it is the major goal to be achieved. The path from bottom is led to Predictive Maintenance (PDM)(F6) and Preventive Maintainace (PVM)(F7) in Level II. Which they drive to Quality of Service (QS)(F8) and Range of Applicability(RA)(F9). They further drive to Telematics (TM)(F10) and Network Effect (NE)(F11) at level IV and leads to Smart Buildings (SB)(F5) at Level V.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

Our research looked into and created links between various factors that affect ESG investing. We used the MICMAC study to identify key aspects that influence ESG investment decisions, and then analysed and generated correlations between those elements in ISM. We were able to identify factor dependency and driving capabilities as a result of the inquiry.

Figure 2 shows how the MICMAC analysis splits the 11 identified IoT enablers into four parts. The first portion is the autonomous zone. It is based on reduced driving and power consumption. There is no autonomous component among the detected components, implying that there is no unfocused factor. The dependent zone is the second part. It is highly reliant and has minimal driving power. Predictive Maintenance (PDM)(F6), Preventive Maintenance (PVM)(F7), Range of Applicability (RA)(F9), and Network Effect (NE) are the major dependent components (F11). The linking zone, the third segment, is unstable and develops with strong driving and dependant powers. Quality of Service (QS)(F8) and Telematics (TM) were both unstable in our investigation (F10). Finally, AIoT (F1), Big Data (BD) (F2), Cloud Computing (CC) (F3), Sustainable Development Goals (SDG) (F4), and Smart Buildings (SB) (F5) are identified as important drivers in the fourth quadrant.

Through this project we can define the need for smart cities, and it also gives us the pathway of how IoT technologies and enablers of smart cities can be applied in urban settings to facilitate cities that work better for their citizens and achieve overall sustainable development. This project shall serve to inform urban planners or scholars as well as Information Commutation Technology experts, and other city stakeholders about the enablers of IoT for smart cities and their sustainable development.

References

1. How do we provide humanity with the affordable energy it needs and stabilize the climate? Whitepaper, Future Stanford| Engineering
2. How good can we get at engineering living matter? Whitepaper, Future Stanford| Engineering
3. How can we ensure that humanity flourishes in the cities of the future? Whitepaper, Future Stanford| Engineering
1. Simon Elias Bibri, “The IoT for smart sustainable cities of the future: An analytical framework for sensor-based big data applications for environmental sustainability”, Sustainable Cities and Society, Volume 38,2018,
2. Y. Kim, Y. Park and G. Song, "Interpretive Structural Modeling in the Adoption of IoT Services," KSII Transactions on Internet and Information Systems, vol. 13, no. 3, pp. 1184-1198, 2019. DOI: 10.3837/tiis.2019.03.004.
3. United Nations, “Population 2030, Demographic Challenges and opportunities for sustainable development planning”, United Nations, New York, 29, July 2015.
4. Gupta, Pooja & Jain, Vijay, “Interpretive Structural Modeling of GIoT enablers” Journal of Information Technology Research,2020
5. Poland, 2015, contribution presented at the CSTD intersessional panel.
6. United States of America Department of Energy, in Cisco, 2010, Cisco smart grid: Substation automation solutions for utility operations, available at http://www.cisco.com/c/en/us/products/collateral/routers/2000-series-connected-grid-routers/white_paper_c11_593673.pdf.
7. Japan, 2015, contribution presented at the CSTD intersessional panel.
8. Deloitte, 2014, Africa is ready to leapfrog the competition through smart cities technology, available ttp://www2.deloitte.com/content/dam/Deloitte/za/Documents/publicsector/ZA_SmartCities_12052014.pd

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

9. Honeywell and Ernst and Young, 2015, Smart buildings make smart cities, available at http://honeywell.com/News/Documents/Smart_Building_Smart_City_WhitePaper_DOWNLOAD.PDF.
10. UNCTAD, 2015, Science, Technology and Innovation for Sustainable Urbanization (New York and Geneva, United Nations publication).
11. United Nations Department of Economic and Social Affairs, 2014, World Urbanization Prospects: The 2014 Revision (New York).
12. Scientific and Technical Advisory Panel, 2014, Sustainable urbanization policy brief, Global Environment Facility; and United Nations Environment Programme, 2015, Cities and climate change,
13. American Planning Association. APA Policy Guide for Planning for Sustainability, New York, April 17, 2000, NY.
14. Vinodh, Sekar & Ramesh, K. & Arun, C. “Application of interpretive structural modelling for analysing the factors influence”,2015
15. Aich, S.; Thakur, A.; Nanda, D.; Tripathy, S.; Kim, H.-C. Factors Affecting ESG towards Impact on Investment: A Structural Approach. *Sustainability* **2021**, *13*, 10868. https://doi.org/10.3390/su131910868
16. D. Gavalas, P. Nicopolitidis, A. Kameas, C. Goumopoulos, Bellavista P., Lambrinos L., Guo B. , Smart Cities: Recent Trends, Methodologies, and Applications, Wireless Communications and Mobile Computing, Vol. 2017, No. 7090963, pp. 2, 2017, Hindawi,
17. Kirimtat, A.; Koyunbaba, B.K.; Chatzikonstantinou, I.; Sariyildiz, S. Review of simulation modeling for shading devices in buildings. Renew. Sustain. Energy Rev. 2016, 53, 23–49.
18. H. Sundmaeker et al., “Vision and Challenges for Realizing the Internet of Things,” Cluster of European Research Projects on the Internet of Things, 2010,
19. A. Gyrard and M. Serrano, “Connected smart cities: interoperability with SEG 3.0 for the internet of things,” in Proceedings of the 30th IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA '16), pp. 796–802, March 2016.
20. Allen, T., and Hoekstra, T. W. (1993). Toward a Definition of Sustainability. Sustainable Ecological Systems: Implementing an Ecological Approach to Land Management. Fort Collins, CO: Rocky Mountain Forest and Range Experiment Station, 98–107.
21. L. Belli et al. IoT-Enabled Smart Sustainable Cities: Challenges and Approaches. Smart Cities. 2020; 3(3):1039-1071.
22. Drepaul, Nicole A. “Sustainable Cities and the Internet of Things (IOT) Technology: IOT Technology Improves the Development of Smart Cities’ Infrastructures and Reduces over-Population Stresses.” Consilience, no. 22 (2020): 39–47.
23. Jeff Desjardins “TECHNOLOGYInfographic: The Anatomy of a Smart City”,Published on January 2, 2019, Visual Capitalist,https://www.visualcapitalist.com/anatomy-smart-city/ 14. Simon Elias Bibr, “The IoT for Smart Sustainable Cities of the Future: An Analytical Framework for Sensor-Based Big Data Applications for Environmental Sustainability”,February 2018,Sustainable Cities and Society 38,DOI: 10.1016/j.scs.2017.12.034i
24. Mamta Pandey, Ratnesh Litoriya and Prateek Pandey,An ISM Approach for Modeling the Issues and Factors of Mobile App Development,,International Journal of Software Engineering and Knowledge EngineeringVol. 28, No. 07, pp. 937-953 (2018)

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

25. Anoop Srivastava,Sanjeev Swami,devinder banwet, Interpretive Structural Modelling of Security System for Better Security Management in Railways,Wiley Publications, 2016
26. Abhinav Pal,Jisha Rajendran(Amity University),Aastha Behl,Application of ISM in the Manufacturing Sector,January 2020
27. S. Vinodh,K. Ramesh & C. S. Arun,Application of interpretive structural modelling for analysing the factors influencing integrated clean sustainable system,Clean Technologies and Environmental Policy
28. Satyabrata Aich,School of Mechanical Engineering Sushanta Tripathy , KIIT University ,An interpretive structural model of green supply chain management in Indian computer and its peripheral industries,
29. Gartner, “Gartner Says 8.4 Billion Connected "Things" Will Be in Use in 2017, Up 31 Percent From 2016,” 2017.
30. Gartner, “Top Trends in the Gartner Hype Cycle for Emerging Technologies, 20,” 2017. 22. Meesun Kim, Hyun Ahn and Kwanghoon Pio Kim, “Process-Aware Internet of Things: A Conceptual Extension of the Internet of Things Framework and Architecture,” KSII Transactions on Internet and Information Systems(TIIS), vol. 10, no. 8, pp. 4008-4022, 2016.
31. Duleba, S., Shimazaki, Y., and Mishina, T., “An analysis on the connections of factors in a public transport system by AHP-ISM,” Transport, vol. 28, no. 4, pp. 404-412, 2013.
32. Attri, R., Dev, N., and Sharma, V., “Interpretive structural modelling (ISM) approach: an overview,” Research Journal of Management Sciences, vol. 2, no. 2, pp. 3-8, 2013.
33. Atzori, L., Iera, A. and Morabito, G., “The internet of things: A survey”, Computer networks, vol. 54, no. 15, pp. 2787-2805, 2010.
34. Sangmin Lee, "Evaluation of Mobile Application in User’s Perspective: Case of P2P Lending Apps in FinTech Industry," KSII Transactions on Internet and Information Systems, vol. 11, no. 2, pp. 1105-1117, 2017.
35. Daebum Ji, Jeongil Choi and Yonghee Kim, “An Exploratory Study on Customer-oriented Attributes for the Revitalization of Digital Healthcare Service :Using Interpretive Structural Modeling,” Journal of Information Technology Services, vol. 17, no. 1, pp. 105-119, 2018.
36. Seul-Ki Choi, Chung-Huang Yang and Jin Kwak, "System Hardening and Security Monitoring for IoT Devices to Mitigate IoT Security Vulnerabilities and Threats," KSII Transactions on Internet and Information Systems, vol. 12, no. 2, pp. 906-918, 2018.
37. Ju, D. Y. and Kim, J. K., “Hyper connectivity era, Activation plan of creative convergence of IoT,” KIET, 2014.
38. Kumar, D. T., Palaniappan, M., Kannan, D., and Shankar, K. M., “Analyzing the CSR issues behind the supplier selection process using ISM approach,” Resources, Conservation and Recycling, vol. 92, pp. 268-278, 2014.
39. Singh, M. D. and Kant, R., “Knowledge management barriers: An interpretive structural modeling approach,” in Proc. of 2007 IEEE International Conference on Industrial Engineering and Engineering Management, pp. 2091-2095, 2007.
40. Wong, J. K. W. and Leung, J. K. L., “Modelling factors influencing the adoption of smart-home technologies,” Facilities, vol. 34, no. 13-14, pp. 906-923, 2016.
41. Mandal, A. and Deshmukh, S. G., “Vendor selection using interpretive structural management,” International Journal of Operations & Production Management, vol. 14, no. 6, pp. 52-59, 1994. 34. European Commission, “Communication from the Commission to the European Parliament, the

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 9 May 2022 doi:10.20944/preprints202205.0102.v1

Council, the European Economic and Social Committee and the Committee of the Regions - Internet of Things : an action plan for Europe," 2009.

42. Raj, T., Shankar, R., and Suhaib, M., "An ISM approach for modelling the enablers of flexible manufacturing system: the case for India," International Journal of Production Research, vol. 46, no. 24, pp. 6883-6912, 2008.
43. Sharma, S., Shimp, T. A., and Shin, J., "Consumer ethnocentrism: A test of antecedents and moderators," Journal of the academy of marketing science, vol. 23, no. 1, pp. 26-37, 1994. 37. Gothwal, S., and Raj, T., "Analyzing the factors affecting the flexibility in FMS using weighted interpretive structural modeling (WISM) approach," International Journal of System Assurance Engineering and Management, vol. 8, no. 2, pp. 408-422, 2017.
44. Yonghee Kim1 , Youngju Park1 and Gwangsuk Song,"Interpretive Structural Modeling in the Adoption of IoT Services",College of Business Administration, Soongsil University Seoul, 06978, South Korea,published March 31 2019,KSII TRANSACTIONS ON INTERNET AND INFORMATION SYSTEMS VOL. 13, NO. 3, Mar. 2019 1184 Copyright ⓒ 2019 KSII
45. Arshad, R., Zahoor, S., Shah, M. A., Wahid, A., & Yu, H. (2017). Green IoT: An investigation on energy saving practices for 2020 and beyond. IEEE Access, 5, 15667–15681. doi:10.1109/ACCESS.2017.2686092
46. Attri, R., Dev, N., & Sharma, V. (2013). Interpretive structural modeling (ISM) approach: An overview. Research Journal of Management Sciences, 2319, 1171. Sierra Wireless. (n.d.). Building Smart Cities. Retrieved from https://www.sierrawireless.com/applications/ smart-cities/ CISCO.(2008). The Internet of Things, Infographic. Retrieved from http://blogs.cisco.com/news/the-internetof-things-infographic Dev,
47. N., Kachhwaha, S. S., & Attri, R. (2013). GTA-based framework for evaluating the role of design parameters in cogeneration cycle power plant efficiency. Ain Shams Engineering Journal, 4(2), 273–284. Footprint Network. (n.d.). Energy Awareness Program. Retrieved from https://www.footprintnetwork.org/ourwork/ecological-footprint/Energy-Awareness-Program
48. Antonio Cilfone 1,2 , Luca Davoli 1,2 , Gianluigi Ferrari 1,2 , Paolo Adorni 3 , Francesco Di Nocera 3 , Alessandro Dall'Olio 3 , Cristina Pellegrini 4 , Marco Mordacci 4 and Enzo Bertolotti "IoT-Enabled Smart Sustainable Cities: Challenges and Approaches Laura Belli",18 September 2020

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

A STUDY ON CLOUD MANUFACTURING

Dr. Ak Madan[*1], Bharath Gohrani[*2], Akshat Singh[*3], Adarsh Pillai[*4]

[*1,2,3,4]Department Of Mechanical Engineering, Delhi Technological University, Delhi, 110042, India.

ABSTRACT

Cloud computing has shown enormous possibility of accelerating growth across all industries, leveraging the concept of utilizing resources remotely. Enterprises and users of cloud systems pay minimal charges to get access to a vast pool of high-performance distributed resources that are delivered to them via the internet. Cloud Service Providers maintain such enormous volume of resources across geographically distributed regions and develop systems to manage them and improve their reliability.

However, developments in the domain of cloud computing are restricted to the field of IT only as of now in industry. This paper aims to expand this idea into the domain of manufacturing and study how the same analogy can be utilized to provide manufacturing, development and design services via the internet to users through a system of distributed infrastructure and allow users to only pay for resources they utilize, which allows them to not invest separately in setting up local manufacturing units and improves efficiency and reduces costs in the long run. Here we first understand the working of cloud systems in detail, see what requirements and functionalities are expected of such systems and finally dive into the technicalities of how such concepts and ideas can be expanded into the manufacturing industry.

I. INTRODUCTION

A steady rise in competitiveness, customer demands, and resource utilization has fueled enterprises to look towards methods of raising productivity and lowering costs while ensuring optimum resource and manpower utilization. A very strong scope of growth and improvement has been observed in the technology sector with the advent of cloud services. This has been one of the 'hot topics' of the 21st century alongside intelligent systems, IOT, Industry 4.0 amongst others. The steadfast growth of several MNCs in the domain of technology such as Google, Amazon, Microsoft and IBM have invested and developed in the concept of creating a business model around selling computing services, hosted online for their clients, known as cloud computing. The primary goal of such an ecosystem is to provide access on-demand to compute-resources such as storage, servers, networks and software in a distributed environment. Users of cloud systems must be able to purchase these as per their needs requiring little management and intervention from their providers alongside having the benefits of scalability and secure access and control.

In cloud computing, the basic model is abstracted into three layers depending on the type of resources that are hosted and provided to users. These are as follows:

The first is known as Infrastructure as a Service (IaaS). In this layer, companies outsource their distributed compute capabilities and hardware as a service for clients to pay and utilize. Typical examples in this category include storage, servers, virtual machines and processing. The physical hard-ware for these is managed by the Cloud Service Provider and may be located across different regions over the world. Provisioning such resources is possible across many locations with allocation being done as per requirement, much of which is automated. The final user may control these resources as a cluster. The target role at this layer is of the administrator who may monitor, deploy these resources as per need and control access within an organization employing cloud facilities. The second layer in the model rests on top of the physical hardware and has more to do with developers and building software. Alongside some hardware level resources, PaaS provides middleware, which is the software layer enabling the flow of data and communication amongst any OS and the applications running on it, software development tools, Database Management Systems and application frameworks, that can help developers in the SDLC process. An added advantage of PaaS is that users get access to communities to build on ideas, debug and collaborate.

The upper layer is Software as a Service, commonly known as an application layer. In this model, software services such as Enterprise Resource Planning (Software used by enterprises in management and operations), Customer Relationship Management and billing control. There softwares are hosted over the internet, so users do not have to install additional software on their system.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

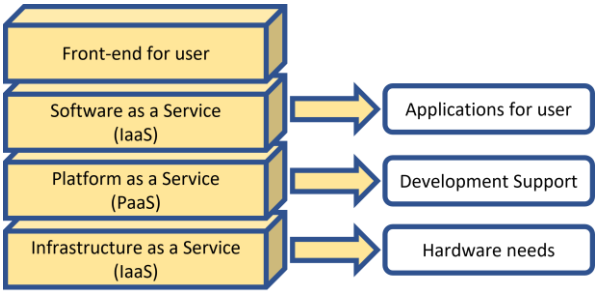


Fig 1: Cloud Service Model

These three components give rise to cloud platforms as we know them today. Although successful implementation of this idea has been seen in the IT sector, more industries are shifting to the use of such concepts to reduce costs, deploy their services efficiently and transform systems and structure. The goal of this research is to introduce the idea of cloud services, but in the manufacturing sector. For long, development in the field of manufacturing has been restricted improving the automatability and output of manufacturing tools and equipment, with little regard to how these can be integrated using IT to expand their scope to a larger pool of users rather than its existing use at the micro-level such as a single manufacturing enterprise or unit. As machine tools become more advanced with the help of CAD/CAM &IOT, the goal is to harness their potential at a geographically larger level and develop on the same cloud concept we have introduced earlier.

II. DETAILED CHARACTERISTICS OF CLOUDSYSTEMS

To understand how manufacturing may be provisioned as a cloud service, it is essential to develop first, an understanding of the architectural requirements and characteristics of an ideal cloud system. It is here that we shall also introduce some common concepts and terminology relating to cloud systems. We can study the requirements of cloud systems from three perspectives, that of the CSP (Cloud Service Provider) itself, enterprises or firms that use cloud resources and services, and single end-users who are tasked with hands-on use.

2.1. Provider Requirements

Beginning with the CSP, deploying the SaaS model is concerned with integration of applications, scalability, fault tolerance and configurability. Monitoring and control of different frameworks used for development is also crucial. PaaS, being concerned with the software development life cycle, has more to do with interoperability of development tools across differ-ent cloud platforms and languages and providing developers with environments for developing, testing and deploying applications. Provision of necessary infrastructure such as stor- age, servers and databases are also crucial. IaaS, being one of the key components due to virtualization of IT resources has to look into providing infrastructure that is self-sustaining and can accommodate many users with complete segregation. How the resources must be provisioned and managed across the platform is another architectural requirement. CSPs must build their systems using standard protocols and frameworks that allow for easy migration of applications and data from one CSP to the other or simultaneous use of multiple CSPs for business applications.

While discussing provider requirements, the term QoS refers to quality of service, which serves as a guarantee between users and providers for performance, security, depend ability and reliability. To implement this, management systems that monitor storage, networking and VMs must be transparent. When IT resources are managed at such a large scale, the chances of failure are large, necessitating fault tolerance which allows systems to continue operation and isolate the causes and assigned users of these resources. Diagnosis, by classification and clustering of resources failing can help in prevention of outages.

The fundamental technology that powers cloud computing is virtualization, which implies the abstraction of resources from the physical level to the logical level. Virtualization allows a single physical resource to be mapped to multiple virtual resources and can be done for servers, storage and networks which allow them to be shared across multiple users or organizations. Demand for software and applications powered by the cloud is always subject to a large variety of fluctuations. The ability of the CSP to handle the provision- ing of cloud resources efficiently to cater to this demand is known as scalability. Scalability can be of two types; horizon-tal scalability refers to the ability of your cloud to efficiently provision all resources currently in use to cater to 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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

load, this is known as load balancing. It avoids a large burden on a single resource and provisions more of the same resource rather than increasing the resource size. Vertical scaling deals with the opposite and adds more computing power to the existing resource to deal with larger application needs. System architecture must accommodate scaling along both axis in a cost and compute efficient manner.

2.2. Enterprise Requirements

For organizations and companies that wish to adopt the cloud,architectural and operational requirements can be laid down as Enterprise requirements that can vary from security to data governance and business process management. Some of these are discussed here. While subscribing to cloud resources, en- terprises can choose amongst 3 deployment models, Public, Private and Hybrid. Public cloud is the most popular model, with cloud services being delivered via the internet with low-cost subscription tier. Here, cloud resources at a single center may be shared with multiple organizations which may cause issues with security and technical control. A private cloud solution refers to dedicated and isolated cloud resources that are provided via secure networks only to a single organization. These resources may be managed by the CSP or any third-party vendors. For the organization, this means greater visibility and control over the resources and greater flexibility to meet any requirements, along with added security. Combining both private and public clouds give us the hybrid cloud, wherein resources are integrated, and applications andloads can share between both depending on factors such as security and performance needs, cost and scalability. A good example of hybrid cloud would be for an enterprise to host its service on the private cloud and switch to public only to accommodate spikes in traffic.

We can discuss the concept of cloud economics, which elaborate the cost benefits of shifting to the cloud for enterprises. Users only pay for the resources they utilize and thus do not need any upfront investments. Besides, cloud providers at any time do not have to deploy their full capacity, since peak demands for each enterprise vary. Cloud providers haveaccess to a large scale of resources and thus have to incur lower costs and may also implement parallel processing efficiently, reducing costs for enterprises. Having cloud centers at dispersed locations allow for redundancy and backup, thus reducing latency and improving speed for enterprises.

Data governance is a critical issue for enterprises and has to do with geographical and political issues that are faced when data is moved across organizational boundaries in situations where it is no longer feasible for it to stay within one provider owing to changes in laws or insolvency. While migrating data from one CSP to another, a challenging is- sue is that of data lock-in, wherein it becomes a hassle to move programs and data across due to lack of interoperability amongst providers. The solution to this is reliance on open protocols and procedures but its implementation is still in grey space, thus leaving enterprises subject to monopolies of cloud services. Lastly, the classical concept of Business Process Management, which helps maintain business structure, security and consistency across the organization can beextended to the cloud as well. The SaaS model is integrated with BPM applications to create management systems for customer interaction, workforce performance and resource planning.

2.3. User Requirements

Lastly, while adopting any cloud system, consideration of user requirements is a must. These include concepts of user privacy, Service Level Agreements and also overall user ex-perience. In the context of cloud, privacy is a challenging issue to tackle. A large amount of user data is to be stored in mega data centers across the globe, susceptible to exposure. Even applications such as social media and email are pow- ered by cloud, making users vulnerable. To deal with this, cloud platforms are reliant on technology such as data com- pressing and encryption, virtual LANs and network boxes that help in failsafe communication. CSPs offer their services to users with the help of contracts known as Service Level Agreements, which lay down the terms of billing and usage and enforces delivery of promised compute resources.

SLAs also lay down the T&Cs for when any clauses of the agreement are violated and dictate the mechanism of compensation to the user. In terms of user experience, the goal of the cloud platform is to learn effectively, the needs and behavior of the end-user which can be used to improve the service in terms of desirability, productivity and usability.One example of the following could be to create a clutter- free UI and UX which allows users to view and provision resources and also to monitor any workflows easily. An ideal UX design development requires in-depth understanding of concepts such as human-computer interaction,

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

ergonomics and usability engineering in the product life cycle.

III. CLOUD IN THE CONTEXT OFMANUFACTURING

Now that we have studied the fundamentals of Cloud systems, it is time to delve into how the same principles and concepts can be applied in the context of manufacturing suchthat benefits of cloud adoption such as agility, scalability and reliability can be reaped. CIOs of a large number of mainstream firms are beginning to consider the possibilityof building strategic business models, wherein cloud can re- place some conventional practices. In this section, we'll takea closer look at one possible method of cloud adoption in caseof manufacturing, that is how some processes in the manufacturing life cycle may be replaced with cloud technologies.Later on, we can take up the more challenging issue of cloud manufacturing, which is the manufacturing analogue of cloud computing.

In earlier sections, we highlighted three key characteristics of cloud systems, one being pay-as-you-go, i.e. users payonly for the services they actually utilize, the second being scalability, i.e. the ability to immediately increase and de- crease resources provisioned as per demand fluctuations and the third being flexibility in deploying and customizing solutions. Adopting cloud computing in terms of manufacturing involves deploying cloud in areas of IT such as HR, Customer Relationship Management and Enterprise Resource Planning. These applications usually fall under the domain of Business Process Management, and the cloud layer involved is usually SaaS or PaaS.

To understand this idea better, we can take a case-study or the real-life example of Elkay Manufacturing Company, an American-owned firm which provides services such as plumbing, interior development and manufactures commercial products such as kitchen and bathroom equipment. Elkaydepicts an example of how cloud services may be applied to improve organization and streamline business activities. With conventional solutions, IT support was having to deal witha lot of manual sales organization and support, which meant dealing with outdated systems and low-level support. Any changes needed in a business process or application would conventionally have to be implemented, customized and inte- grated in a hardcore manner to suit the evolution and would involve support from outside vendors on a contract basis, as well as higher time consumption. With the introduction of cloud, process change can happen seamlessly and faster. Furthermore, a workload shift has been observed amongst theIT staff at the firm to higher level duties along with expertiseand skill development.

The entire migration from conventional systems to the cloudfor Elkay was handled by an implementation and integration consulting firm, ArganoCSS, to develop and end-to-end on- premises Oracle cloud solution, deploying the hybrid model that we have discussed earlier. Some of the key factors and services opted for in the process have been elaborated here for a detailed understanding.

- **Procurement Cloud:** Oracle's procurement cloud is an integrated source-to-settle solution that allows for strategic stock sourcing, development of supplier rela- tionship management and reducing overall risk while buying. A well-developed UI and UX improves shop- ping experience and streamlines payments. Any sup- ply chain requirements are fulfilled with resiliency andquality customer service. Cloud intelligent services canalso be deployed here to analyze and improve spending, while gaining insights into spend patterns and identify-ing shifts in market dynamics. Supplier Relationship Management also helps to verify suppliers and monitordevelopments to reduce risk and disruption.



Fig 2: Procurement Cloud

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

- **Enterprise Performance Management:** As organizations grow larger, essential activities such as workforce planning, capital asset planning and budgeting become challenging if implemented conventionally. EPM solutions allow companies to align functions of planning, budgeting and forecasting across all lines of a business. As the volume of data increases, operational planning needs to be segregated from corporate system administration. This is where EPM allows for creation of a unique architecture that aids in creating forecasts, plansand budgets.
- **E-Business Suite:** EBS is an integrated set of business applications which aids in automating CRM, ERPand supply chain management within organizations. contains application modules that support different aspects of CRM, ERP and SCM operations, including customer service management, financial management,human capital management, project portfolio management, procurement, and value chain planning and execution. Oracle also has integrated its business intelligence applications with E-Business Suite to provide EBS users with BI and analytics capabilities. In this context, moving EBS to the cloud helps improve business agility, manage growth, reduce time and costs andimprove staff productivity, utilizing cloud resources rather than private data centers.

Besides Elkay, another example of cloud adoption could be towards enterprise collaboration at a broader scale. Within the organization as well, different departments can collab- orate amongst each other like demand planning and supply chain, or sales teams with others for foreseeing opportunities.Cloud technology in this case can provide a collaborative environment that gives agility and transparency.

IV. CLOUD MANUFACTURING

In some context, the idea of cloud manufacturing can be viewed in synonymity with that of internet-based manufacturing or distributed manufacturing. Distributed Manufacturing (DM) is generally a process of manufacturing whose endpurpose is to reduce the distance between the end user of theproduct and its location of manufacture, i.e. production as close to the source of need as possible. A large number of firms around the globe have begun exploiting this idea, by making use of a large pool of manufacturing resources and networks around the world. Such a decentralized method of production and distribution is powered by online technology and the Internet of Things (IoT), connecting geographically dispersed facilities.

The benefits of such as a system are many, the most important being cutting down in shipping costs by assigning production tasks to local manufacturers in the network. This also makes the process more sustainable, leveraging local expertise and smaller volumes of production. For enterprises having smaller resources and scale, distributed manufacturing helps expose them to the process of globalization and outsourcing to reduce capital investment and overhead expenses. The crucial difference between DM and traditional manufacturing is that conventional systems employ a single large facility, shipping product to the customer, whereas in DM, several decentralized units across a network are used to reduce geographical disparity between users and firms. Inventory management is one more advantage of DM, allowinglarge firms to ease their inventories with excess goods and save capital investments. The idea of DM can be seen in figure 3.

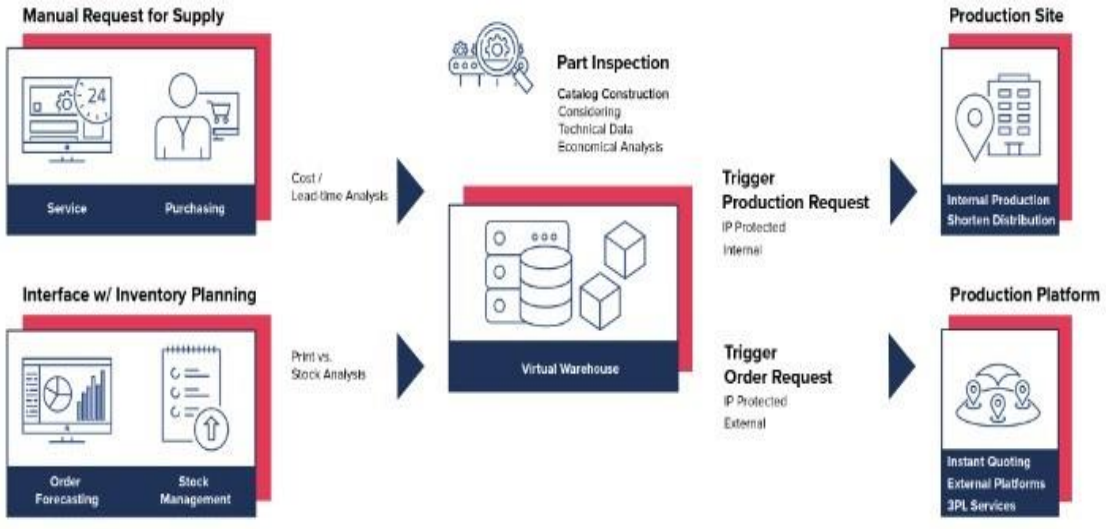


Fig 3: Concept of Distributed Manufacturing

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

One drawback of such systems, however, is that there is little to no coordination amongst the resource serviceproviders and the resource service demanders throughout theprocess. This is in stark contrast to cloud computing concepts we have seen earlier. Besides, what is lacking in the DM system is the centralized operation and management of distributed resources. Without these characteristics, providing seamless and high-quality transaction of resources is not possible. Owing to such causes, the adoption of DM has not been so effective.

It is now that we introduce the concept of Cloud Manufacturing, which does offer a more attractive solution compared to DM. Some of the characteristics we have seen in cloud computing such as Pay-as-you-go, virtualization, content outsourcing and so on have their manufacturing analoguesin Cloud Manufacturing. It encompasses services such as networked manufacturing, Manufacturing grid, virtual man- ufacturing, agile manufacturing and IoT, along with cloud computing. The goal of Cloud Manufacturing is to encapsu- late distributed manufacturing resources and provide them as a bundle of cloud services, which can be centrally managed. Users of CM systems can opt for services such as product de-sign, testing, manufacturing, management and other activities in the product life cycle. We can design and develop a cloud manufacturing solution in 4 layers as shown in figure 4. Eachdomain depicts its user of concern.

4.1. Manufacturing Resource Layer

At the lowest level we have the Manufacturing Resource Layer and the Virtual Service Layer, both of which are managed and controlled by the cloud manufacturing ser-vice provider. The Manufacturing Resource Layer comprisesof two components,

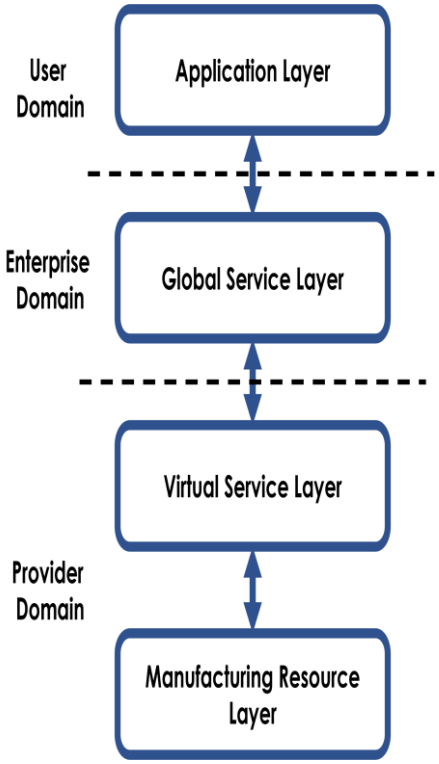


Fig 4: Cloud Manufacturing High-Level Architecture

- Physical Hardware and resources such as machine tools, computers, servers and raw materials, and,
- Manufacturing Capabilities, which are intangible goods such as knowledge and expertise in the areas of design- ing, testing, simulations, experimentation and manage-ment, besides others.

When such resources are provided over the cloud, we can draw their comparison with cloud computing as they can be provided in the form of IaaS and SaaS.

4.2. Virtual Service Layer

The second component of this section comprises of the virtual service layer. Having distributed manufacturing resources initself is not enough. We need to find ways in which we can identify these across a network, convert them into virtual resources and package them to be provisioned as cloud ser- vices. This is however, a

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

challenging task to accomplish. Various technologies can be used for identifying and tagging manufacturing resources such as Wireless Sensor Networks,RFID systems, Cyber Physical Systems and IoT, besides others.

We can study the working of two such systems as an exam- ple here. We consider the RFID system first, which consists of two components, a RFID tag for the object we wish to detect and a RFID reader which emits radio waves. The tag then receives these waves and uses their energy to transmita feedback signal which helps in its detection. RFID tags are usually passive and do not carry their own energy source.The RFID reader consists of a signal generator and a sig- nal receiver, along with a microcontroller to process data. The RFID tag consists of a transponder, to receive and send signals, and a rectifier circuit for extracting energy from in- coming radio waves to power its controller. Figure 5 shows the construction of a RFID tag and reader.

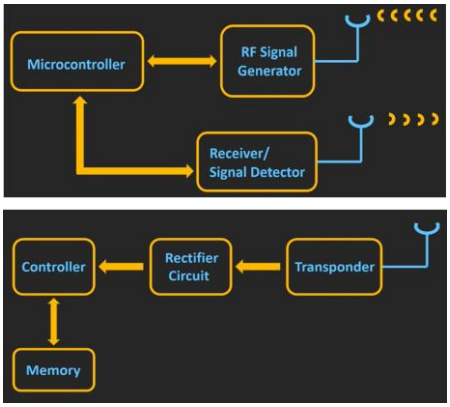


Fig 5: RFID Tag (above) & Reader (below)

We can also study the concept of Wireless Sensor Net- works here. The overall objective of a sensor is to take any physical quantity from its environment and convert it into an observable quantity, that can be used for further analysis or decision making. A sensor network consists of many such sensor nodes, which are lightweight and highly distributed to monitor any system or environment. In most cases, such sensors are distributed in an ad-hoc manner.

We can see the construction of a basic WSN in figure 6.It consists of many such sensor nodes, along with a single base station. The purpose of the base station here is to col- lect data or sensory information from each node and transmitit further. The base stations are connected to the internet through which data collected can be used for mining, processing or analysis.

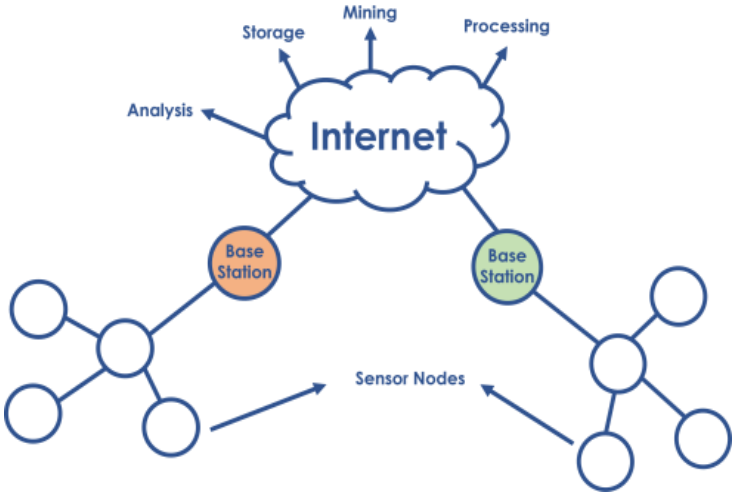


Fig 6: Wireless Sensor Network

When we define the concept of virtualization, we refer to abstraction of physical resources to create logical resources that can be provided to users. How well can we accomplish this, or the quality of virtualization is crucial for ensuring a robust cloud infrastructure. Manufacturing resources, expertise and know-how can be virtualized in the same manneras conventional cloud IT resources. How we can do this, is explained next.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

MTConnect is an example here, of a standard based on openprotocol for data integration. The main purpose of MTConnect is to support and enhance data acquisition capabilities ofmachine tools, but its ability to support a plug-and-play environment for manufacturing, which can be used to remotely interact with the system, has allowed it to be a suitable method of supporting cloud manufacturing. Figure 7 shows the high-level system architecture of MTConnect. At the coreof the system lies the agent, which acts as a bridge between the device, which can be a machine tool and the client application. The agent can receive and store data samples which can be used for further communication.

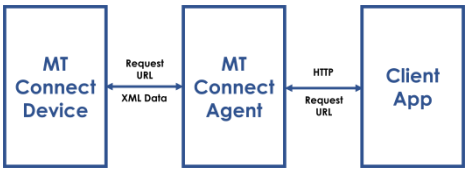


Fig 7: MTConnect Architecture

The underlying protocol that is used for communication in MTConnect is HTTP, used throughout the internet and theWorld Wide Web. This ensures that devices using MTCon- nect are accessible worldwide, through the internet. HTTP also defines the communication standard that can be used to access pages and data, in the form of URLs (Uniform Re- source Locators) as we shall see next.

Communication amongst the agent and the device involves the flow of certain types of data, which include:

- Physical and actual device design data
- Measurement and calibration data
- Real time data from the device

The flow of this data from the device to the agent is accomplished by using XML (eXtensible Markup Language) to encode the data and allow for communication. The advantageof XML here is that its structure is hierarchical, which allowsit to resemble the construction and feature hierarchy of a ma-chine tool. Having such a definition allows for faster access and modification to machine data, with fewer commands. The chart in figure 8 shows an example hierarchy for a ma- chine tool that can be represented with XML in a MTConnectsystem. The next step of the process is to package these manufacturing capabilities and provide them as cloud services. The MTConnect agent used in this system is quite flexible and handles requests with the use of HTTP as we have seen earlier. The concept of URLs is prominent here and URL parameters are used in order to pass various commands to theMTConnect Device. A typical request consists of a URL, anaction and a query, similar to giving a URL address through the web. The URL gives the name of the machine to be used and also a link to any manufacturing resource such as amachining program. The action and query describe in detail what needs to be done with the given resource.

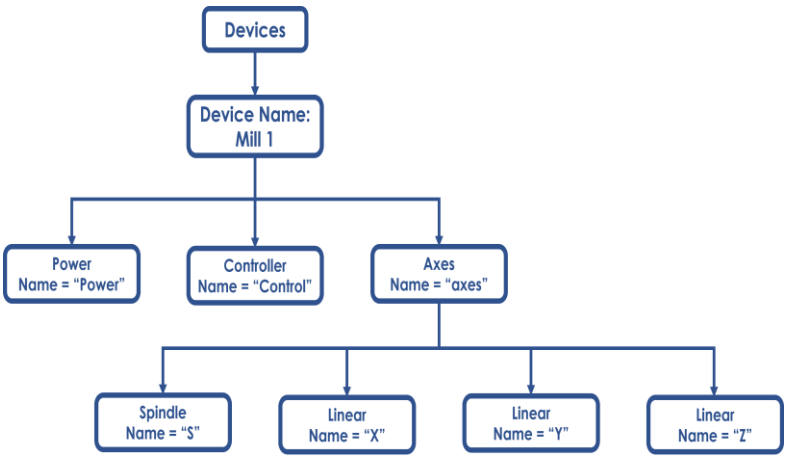


Fig 8: Machine Tool XML hirearchy with MTConnect

Figure 9 shows an example of a factory layout having three machines virtualized and integrated via MTConnect. These are connected to onsite data monitoring and offsite analysis tools, all by using open protocols and technologies. Mapping of manufacturing resources to create virtualized cloud ser- vices can be done in three

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

different ways as discussed below.

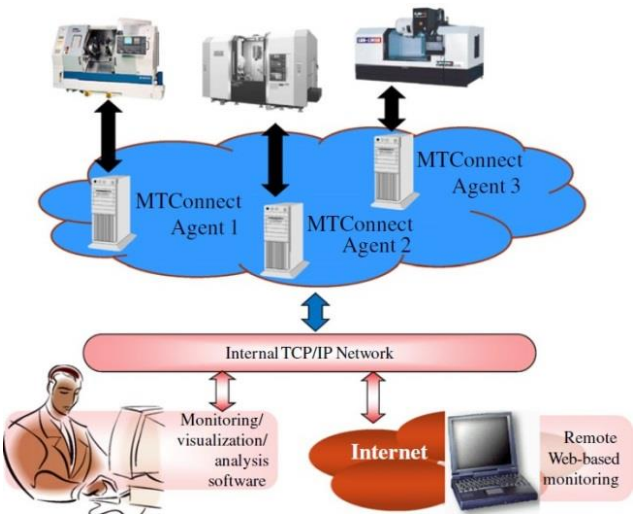


Fig 9: Factory Layout with MTConnect

The first is one-to-one mapping, the simplest of them all.In this mapping, manufacturing resources that can provide only a single function are mapped to form one cloud service.For example, CAD files extensions that are readable by 3D printing software or CAE applications are only readable by them. Likewise, EDM tools can only be used for the EDM process and thus can be used to form a single cloud service.

In many-to-one mapping, multiple manufacturing resources are combined in order to create a more generalized, powerful cloud service. For example, software that can separately provide meshing and simulation abilities can be combinedto form a generic simulation service, available as a single service over the cloud.

One-to-many mapping is rather simpler, wherein a single manufacturing resource can be encapsulated and provided asmultiple cloud services to the end user, who shall see separate functions and abilities. For example ANSYS is a very powerful simulation tool with CFD, thermal analysis, electromagnetic analysis and so on. These can be provisioned as standalone cloud services.

4.3. Global Service Layer

Until now we have seen the bottom two layers of the Cloud manufacturing stack, namely the manufacturing resources and hardware and the virtualization layer, wherein these re- sources can be accessed and utilized. Having studied this, there remains still, a need for effectively managing the al- location and distribution of cloud resources and centralizing its operation. This is where the Global Service Layer comes into the picture. Depending on the type, availability of manufacturing resources provided and the requirements of the user, we can have two operation modes at the Global ServiceLayer, complete service and partial service.

When the complete service mode is in place, the full responsibility of the cloud manufacturing operation is taken by the Global Service Layer. The category of manufacturing resources that can be managed to such an extent at the complete service mode are CPUs, RAM, Virtual Machines and so on. The reason for this is that these resources can be dynamically monitored, managed and load-balanced with ease.Any form of application software can also be a part of this, since its running and execution can take place in a distributed computing environment. Besides these, knowledge, human resources and manufacturing capabilities are also managed in complete service mode.

When it comes to providing services like manufacturing hardware, the cloud service provider can only handle the operation partially, wherein some additional input and operation services are provided. This is known as partial service mode.In this mode, the global service layer is responsible for activities like allocating, locating, billing and remote monitoring the process. Those who provide their hardware to the cloud service providers are still responsible for executing the manufacturing tasks and ensuring the quality of product formed.

At the global service layer, a crucial task of cloud manufacturing that has to be implemented to meet the above requirements is of optimum resource allocation. There are several methods that are used to do the same. We

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

shall take one for example here, of the Particle Swarm Optimization technique.

PSO is an algorithm that is bio inspired and tailored to searchfor an optimal solution, given the solution space. One advantage of using this method is that only the objective function is utilized to converge, and it does not depend on any differential form or gradients. This method uses very few hyperparameters. The intuition for PSO was inspired by the social behavior of bird flocking or fish schooling where any member of the group can profit from the experience of other members whenworking on a mutual objective. This algorithm is Population-based, implying that its study involves dealing with a population or a group of particles in a formation or collection, known as swarm. The algorithm is stochastic, implying that solutions are searched for in a random manner or process.

The boilerplate algorithm for the PSO process is demon- strated as follows:

FOR each particle

INITIALIZE particle

Initialize a nested loop and compute the fitness values for each Virtual Machine on each task using the values from the objective functions. Once we have these fitness values, we can rank different virtual machines corresponding to their suitability for each task for every objective function.

We ENDDO FOR each particle Calculate fitness value If fitness value better than pbest, replace END Choose best fitness value as global bestFOR each particle Calculate particle velocityUpdate particle position END.

Here fitness value is a measure of how well or fit the given candidate is to form a solution for the problem. The particlesare moved in the search space according to a simple mathematical formula over the particle's position and velocity. Each particle's movement is influenced by its local best- known position, but is also guided toward the best-known positions in the search-space, which are updated as better positions are found by other particles. This is expected to move the swarm toward the best solutions.

In the context of cloud manufacturing, we use a more complex version of this algorithm, known as the Multi-ObjectiveParticle Swarm Optimization algorithm. In this context, our final objective is to find the suitable virtual machine or manufacturing resource for any given operation. Manufacturing resources are mapped to become virtual machines as we havestudied earlier. We have several different tasks and many virtual machines available to us. Our objective is to find which one is the most suitable.

The three objectives that we have to account for are,

- Minimizing the Task Execution Cost
- Minimizing the Expected Completion Time
- Maximizing the Virtual Machine Processing Power

We start off in a similar manner, with the MOPSO algorithmillustrated in the figure below.

We select a small group of VMs to begin with. For each task on each VM, we calculate the values of the three objective functions and store them in memory. We once again

```
Procedure TASKSCHEDULING (T,VMs)
   for t ϵ T do                                    // iterate all tasks
      for v ϵ VMs  do
            TEC(t,v)← CALCULATETEC(t,v )     // cost of each task
            ECT(t,v)←CALCULATEECT(t,v )      //task execution time
      end for
   end for
   INITIALIZE (Swarm , Velocity , Position, Pbest)
   for i ϵ T do
      for v ϵ VMs do                          // iterate through all particles
            f=COMPUTEFITNESS((VMs,ECT,TEC)
            Pbest (i) =EVALUATE(f)
            UPDATE(Velocity(i), Position(i))
      end for
   end for
   return Pbest
end Procedure
```

Fig 10: MOPSO Algorithm

can create matrices as shown below and using the calculated values, rank the VMs. We then sum up ranks for each cell for all three objective functions and use the best possible option. Once we do that, we update the position and velocity, imply-ing that selecting a new group of VMs that can give a better solution and continue

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

the algorithm until we are satisfied withits convergence.

Table 1: Task Execution Cost (TEC)

The values of TEC					
Task	V1	V2	V3	V4	V5
T1	10	12	8	15	12
T2	15	11	13	12	10
T3	12	14	10	10	8
T4	14	18	16	22	19
T5	18	12	19	17	15
T6	18	16	19	15	14
T7	12	18	22	20	19
T8	24	20	18	14	17

The rank value after sorting					
Task	V1	V2	V3	V4	V5
T1	2	4	1	5	3
T2	5	2	4	3	1
T3	4	5	2	3	1
T4	1	3	2	5	4
T5	4	1	5	3	2
T6	4	3	5	2	1
T7	1	2	5	4	3
T8	5	4	3	1	2

Table 2: Expected Completion Time (ECT)

The values of ECT					
Task	V1	V2	V3	V4	V5
T1	21	24	30	22	25
T2	24	21	19	25	27
T3	18	21	21	26	16
T4	25	20	24	19	21
T5	30	25	23	22	26
T6	32	30	21	25	28
T7	19	17	18	22	23
T8	12	19	21	17	20

The rank value after sorting					
Task	V1	V2	V3	V4	V5
T1	1	3	5	2	4
T2	3	2	1	4	5
T3	2	3	4	5	1
T4	5	2	4	1	3
T5	5	3	2	1	4
T6	5	4	1	2	3
T7	3	1	2	4	5
T8	1	3	5	2	4

Fig 11: TEC & ECT

4.4. User Requirements: Application Layer

The application layer serves to be an interface between the user and the cloud manufacturing services. This layer is composed of complex system modelling tools, computersimulation terminals and product development utilities. Similar to IT cloud systems, the user here can schedule and run a manufacturing job using manufacturing resources to generatevalue.

Table 3: VM Processing

The Values of VM processing					
Task	V1	V2	V3	V4	V5
T1	20	14	12	21	15
T2	21	15	20	24	18
T3	27	18	24	30	25
T4	27	23	21	24	30
T5	29	23	22	28	26
T6	32	26	25	23	22
T7	35	28	21	24	20
T8	21	28	25	23	25

The rank value after sorting					
Task	V1	V2	V3	V4	V5
T1	2	4	5	1	3
T2	2	5	3	1	4
T3	2	5	4	1	3
T4	2	4	5	3	1
T5	1	4	5	2	3
T6	1	2	3	4	5
T7	1	2	4	3	5
T8	5	1	2	4	3

Table 4: VMs Rank Values

Task	V1	V2	V3	V4	V5	Best
T1	2+1+2=5	4+3+4=11	1+5+5=11	5+2+1=8	3+4+3=10	V1
T2	5+3+2=10	2+2+5=9	4+1+3=8	3+4+1=8	1+5+4=10	V3
T3	4+2+2=8	5+3+5=13	2+4+4=10	3+5+1=9	1+1+3=5	V5
T4	1+5+2=8	3+2+4=9	2+4+5=11	3+5+1=9	4+3+1=8	V1
T5	4+5+1=10	1+3+4=8	5+2+5=12	3+1+2=5	2+4+3=9	V4
T6	4+5+1=10	3+4+2=9	5+1+3=9	2+2+4=8	1+3+5=9	V4
T7	1+3+5=9	2+1+2=5	5+2+4=11	4+4+3=11	3+5+5=13	V2
T8	5+1+5=11	4+3+1=8	3+5+2=10	1+2+4=7	2+4+3=9	V4

Fig 12: VM Processing & Total Ranking

In the same manner as to how users are billed for utilizing compute resources and IT infrastructure, cloud manufacturing users are faced with consumption-based billing and metering for utilizing resources in manufacturing such as water, gas and electricity. To improve transparency, activity- based costing is implemented which allows users to view beforehand, how much their activity would cost them in the terms of cloud manufacturing charges.

User privacy in cloud manufacturing systems is a challenging question to deal with. Storage of personal and enterprise sensitive data containing product information and manufacturing expertise susceptible to leaks and failure bring up the need toenforce some confidentiality and security layers such as:

- Compressing data and encrypting it
- Using virtual LANs that can offer secure remote communications
- Network middle boxes, for failsafe communications

A Service Level Agreement, which lays down the detailed terms and conditions along with the methodology in which cloud manufacturing services shall be extended to the user is an important element before utilizing any feature or service from the CSP.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

The entire product life cycle from design to manufacture mustinvolve the use of a single language and protocol, that makes data portability easier and sets a standard across the industry. Many open data formats and APIs have been suggested for the same. One such solution is the use of STEP (Standard forExchange of Product data) which provides a mechanism for describing the product data, independent of any system. This format of organizing data makes it suitable not just for file exchange but also as a basis for implementing , sharing and archiving product data over a cloud manufacturing system.

V. RESEARCH PROTOTYPE ON A CLOUDMANUFACTURING SYSTEM

Until now we have observed a theoretical explanation and demonstration of how a cloud manufacturing service can function and how does its architecture work. We studied a layer-wise construction of a cloud manufacturing system andwhat tools and systems each layer encompasses, from the CSP to the cloud user.

For our next section, we will dive into existing available literature and explain one such attempt and constructing a cloud manufacturing service, making full use of concepts wehave learnt such as cloud computing, IoT and other relevant technology. This prototype has been developed by the Engineering Research Center of Complex Product Advanced Manufacturing Systems, Ministry of Education, Beihang University, Beijing, China.

The prototype in question here encompasses design and simulation-oriented cloud services and focuses on the variouskinds of resources in the design and simulation phase of a product life cycle. The cloud manufacturing solution consistsof four subsystems, which are elaborated below.

5.1. Virtualization Supported Subsystem

The concept of virtualization that we have described earlier is realized this system of the cloud manufacturing platform. Virtualization here includes features such as customization, encapsulation, deployment and testing of virtual machine. The virtualization subsystem consists of two modules, the resources perception system and the virtual access system module. The resources perception subsystem is responsible for the acquisition of resources, their analysis and aggregation for being used as cloud services. Since this module is involved in acquisition, it is deployed on the end of the cloudmanufacturing resource provider. The virtual access sys- tem module takes these distributed resources and maps them into virtual resources. This includes customization, implyingforming of virtual resources as per requirements, monitoringof these VMs and building fault tolerance. The virtualizationsupported system is shown in the figure below.



Fig 13: Virtualization Supported Subsystem

5.2. Cloud Service Management Subsystem

Registration of new cloud service, managing the distribution of services and extending support to the management of cloud services can be provided in this layer. While registering a new cloud service, a standard description model is used to get a description of resource services. Three function modules are present in this management system, semantic publishing of aservice, querying it and modifying it.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

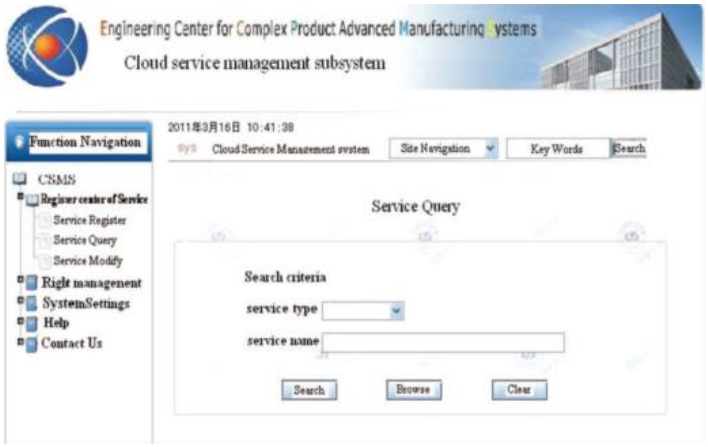


Fig 14: Cloud Service Management Subsystem

5.3. Knowledge Database Management System

This is one of the most crucial components of supporting the cloud manufacturing platform's operation. Here ontology is taken as the form of knowledge representation. this knowledge database system can provide the application interface for other subsystem to realize the seamless connection between knowledge database system and cloud service platform.

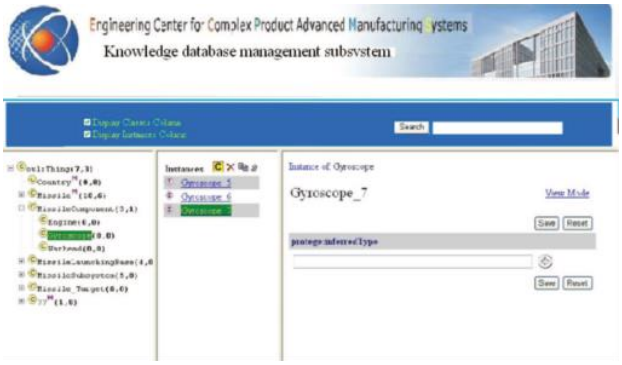


Fig 15: Knowledge Database Management System

5.4. Cloud services oriented complex product design sub-system

This subsystem takes cloud service as a core, provides knowledge management and semantic supporting by knowledge database, provides various application services of life cycle of products to users on-demand through network, thus efficient sharing and collaboration of cloud services are achieved.

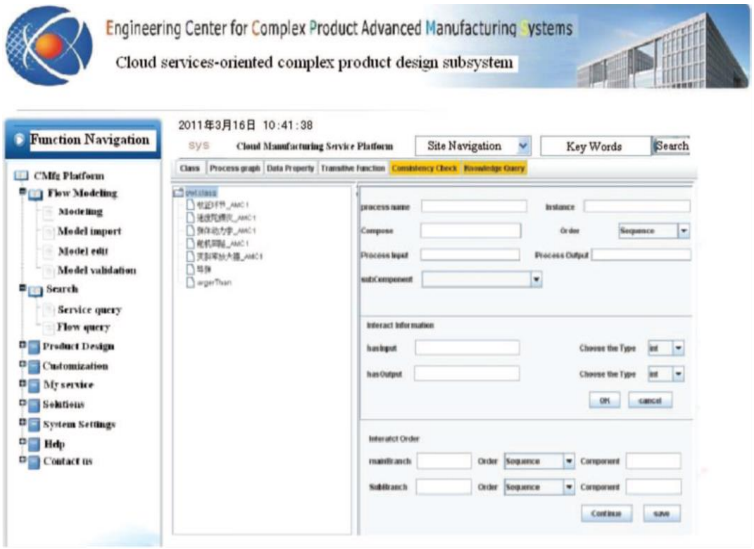


Fig 16: Cloud services oriented complex product design sub-system

VI. CONCLUSION

Throughout the course of this paper, we have covered in detail the scope and current developments in the domain of cloud manufacturing. We first discussed in detail the concepts of Cloud computing in general, including an overall frameworkand architecture. To understand how cloud serves our purpose, we developed and presented some insights into what expectations and requirements are made of cloud systems, from the viewpoints of cloud service providers, enterprises that purchase and utilize these services and users who interactwith these cloud systems.

Once we built a solid foundation on the concepts of cloud, webegan exploring the domain of manufacturing, first looking at how cloud with IT focused infrastructure can assist manufacturing. Our section of cloud manufacturing explained a full- fledged manufacturing-focused cloud system showing how manufacturing resources can be virtualized and packaged as cloud provisioned services in a distributed environment and how such a structure would function, including algorithms for resource provisioning. Finally we explored available literature to showcase a prototype developed for such a cloud manufacturing system. This area holds tremendous potential for creating a disruptive, innovative growth in the domain of manufacturing. A greater degree of cooperation and focusis required both from industry and academia to develop this concept further.

VII. REFERENCES

[1] Lei Ren et al. “Cloud manufacturing: key characteristics and applications”. In: International journalof computer integrated manufacturing 30.6 (2017), pp. 501–515.

[2] Lin Zhang et al. “Cloud manufacturing: a new manufacturing paradigm”. In: Enterprise Information Systems 8.2 (2014), pp. 167–187.

[3] Gui Yun Tian, Guofu Yin, and David Taylor. “Internet- based manufacturing: A review and a new infrastructure for distributed intelligent manufacturing”. In: Journal of intelligent manufacturing 13.5 (2002), pp. 323–338.

[4] Xun Xu. “From cloud computing to cloud manufacturing”. In: Robotics and computer-integrated manufacturing 28.1 (2012), pp. 75–86.

[5] Entisar S Alkayal, Nicholas R Jennings, and Maysoon F Abulkhair. “Efficient task scheduling multi-objectiveparticle swarm optimization in cloud computing”. In: (2016), pp. 17–24.

[6] K Cheng, DK Harrison, and PY Pan. “Implementationof agile manufacturing—an AI and Internet based approach”. In: Journal of Materials Processing Technology 76.1-3 (1998), pp. 96–101.

[7] Bhaskar Prasad Rimal et al. “Architectural requirements for cloud computing systems: an enterprise cloud approach”. In: Journal of Grid Computing 9.1 (2011), pp. 3–26.

[8] James Kennedy and Russell Eberhart. “Particle swarmoptimization”. In: 4 (1995), pp. 1942–1948.

[9] Matthew NO Sadiku, Sarhan M Musa, and Omonowo D Momoh. “Cloud computing: opportunities and challenges”. In: IEEE potentials 33.1 (2014), pp. 34–36.

[10] Lizhe Wang et al. “Cloud computing: a perspective study”. In: New generation computing 28.2 (2010), pp. 137–146.

[11] Lei Ren et al. “Cloud manufacturing: from conceptto practice”. In: Enterprise Information Systems 9.2 (2015), pp. 186–209.

[12] Dazhong Wu et al. “Cloud manufacturing: Strategic vi- sion and state-of-the-art”. In: Journal of Manufacturing Systems 32.4 (2013), pp. 564–579.

Archives of Computational Methods in Engineering
https://doi.org/10.1007/s11831-022-09744-5

REVIEW ARTICLE

A Systematic Review on Breast Cancer Detection Using Deep Learning Techniques

Kamakshi Rautela[1] · Dinesh Kumar[1] · Vijay Kumar[2]

Received: 16 September 2021 / Accepted: 11 April 2022
© The Author(s) under exclusive licence to International Center for Numerical Methods in Engineering (CIMNE) 2022

Abstract
Breast cancer is a common health problem in women, with one out of eight women dying from breast cancer. Many women ignore the need for breast cancer diagnosis as the treatment is not secure due to the exposure of radioactive rays. The breast cancer screening techniques suffer from non-invasive, unsafe radiations, and specificity of diagnosis of tumor in the breast. The deep learning techniques are widely used in medical imaging. This paper aims to provide a detailed survey dealing with the screening techniques for breast cancer with pros and cons. The applicability of deep learning techniques in breast cancer detection is studied. The performance measures and datasets for breast cancer are also investigated. The future research directions associated with breast cancer are studied. The primary aim is to provide a comprehensive study in this field and to help motivate the innovative researchers.

1 Introduction

Breast cancer is categorized among the most frequently reported cancers in the World. It has been reported in both males and females. However, its frequency with females is far beyond the comparison. In 2018, it is estimated that 6,27,000 women died due to breast cancer, which is approximately 15% of all cancer deaths among women [1, 2]. The early detection of breast cancer may help the patient to be recovered in time. However, it is advisable not to go for frequent breast cancer screening due to lack of convenience and discomfort with traditional examinations such as mammograms. It is reported in literature that 2,68,600 females suffered from breast cancer out of 2,71,270 cases. Breast cancer alone accounts for 30% of all new cancer diagnoses in women. The estimated number of deaths in both cases is 42,260. However, the death rate (number of deaths) of women (i.e., 41,760) is much higher than men (i.e., 500). The early screening and treatment of this cancer can be helpful to decrease the mortality rate.

According to National Centre for Disease Informatics and Research (NCDIR), the estimated breast cancer cases and mortality in females in India are shown in Fig. 1. It is observed from Fig. 1 that the estimated number of breast cancer cases in 2016 was 1,56,423 and increased to 1,78,361 in 2020. The estimated number of deaths in 2016 was 80,973. 90,408 women died due to breast cancer in 2020. Figure 2 shows the number of new cancer cases in India in 2020. According to National Agency for Research on Cancer, the number of new breast cancer cases is 1,78,361, which is approximately 26% of all cancer-related cases registered in 2020.

Male breast cancer represents 1% only of all breast cancer cases [4]. The frequency of breast cancer in transgender individuals, as well as the impact of gender-affirming hormonal treatment (GAHT) on the risk of breast cancer, remains largely unexplored. It is less clear however, what risk breast cancer poses to the transgender individual and how, if at all, physicians should screen these patients. Reports of transgender men breast cancer have been mentioned in the medical literature [5]. Number on the incidences of breast cancer in trans women receiving GAHT remains vague. As of 2018, two population-based studies assessed the breast cancer risk attributable to GAHT. Both studies were limited by small number of breast cancer cases and a lack of genetic risk stratification [6, 7]. In the trans man case, ductal carcinoma in-situ (DCIS) was diagnosed in the course of chest reconstruction surgery. To maintain masculinization,

✉ Vijay Kumar
vijaykumarchahar@gmail.com

[1] Delhi Technological University, New Delhi, India

[2] National Institute of Technology, Hamirpur, Himachal Pradesh, India

Published online: 25 April 2022

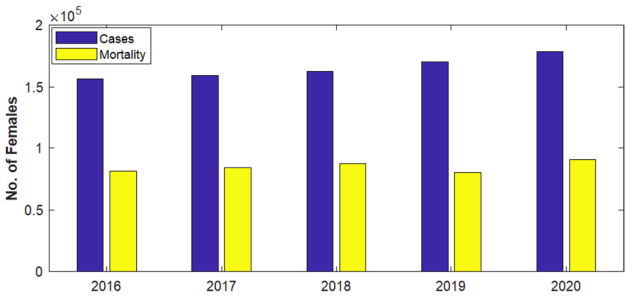


Fig. 1 Breast cancer cases and mortality in females in India [3]

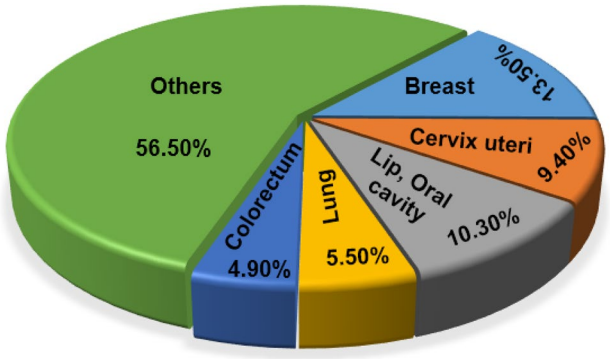


Fig. 2 Number of new cases of cancer found in India during 2020

low dose transdermal application of testosterone may be applied as these doses may minimize the amount of circulating testosterone and thus avoid unnecessary aromatization to estradiol [8]. In [7], authors suggested the risk of breast cancer in transgender people is lower than cisgender women, however, it is comparable to the risk in cisgender men. They also concluded that the overall risk of breast cancer in transgender people remains low. Therefore, it seems sufficient for transgender people using hormone treatment to follow screening guidelines as for cisgender people. Table 1 shows the cancer data statistics 2020 for India. This data is obtained from National Centre for Disease Informatics and Research (NCDIR). According to the data, Breast Cancer is amongst the top 5 most frequent cancers in India.

Normally, patients with breast tumors undergo multiple different examinations including B-ultrasonography, Mammography, Computed Tomography (CT), and Nuclear Magnetic Resonance Imaging (MRI) [9]. Mammography is the main method used for screening breast cancer. Mammography is the only imaging test that reduces breast cancer mortality [10–12]. Mammogram does not prevent cancer. However, the early detection of cancer can be possible through mammography [13]. The sensitivity of mammography is estimated between the range of 77% and 95%. The specificity estimated through the mammography lies in the range of 92% to 97% [14]. However, mammography is suboptimal in breasts with dense tissue [15]. Due to this, approximately 38% of tumors are missed or misdiagnosed [16]. Another drawback of mammography is subject discomfort and radiation exposure. The interpretation of mammograms is a time-consuming and error-prone task [17].

1.1 Motivation

From the literature, it is observed that while substantial strides have been made, the prevalence of cancer tends to increase. For women globally, breast cancer is the most prevalent disease. A woman's chance of breast cancer today is one in eight [18]. Altering is a co-morbidity of breast cancer that can promote the development of breast cancer due to aging-related transcriptome changes [19, 20]. However, the age of females diagnosed with cancer is shifting from +50 years of age to 40 years of age or less [21]. Early identification of breast cancer improves prognoses according to the World Health Organization (WHO) [22]. Mammography is the main method for breast cancer detection. The sensitivity of mammography decreases with increase in thickness of breast [23, 24]. Hence, it is important to understand breast cancer detection techniques so that the

Table 1 Statistics of Cancer Data in India during 2020

	Male	Female	Total
Population	717,100,976	662,903,415	1,380,004,378
Number of cancer cases	646,030	678,383	1,324,413
Number of cancer deaths	438,297	413,381	851,678
Age-standardized incidence rate (World)	95.7	99.3	97.1
Age-standardized mortality rate (World)	65.4	61	63.1
5-year prevalent cases	1,208,835	1,511,416	2,720,251
Top 5 most frequent cancers excluding non-melanoma skin cancer (ranked by cases)	Lip, oral cavity Lung Stomach, Colorectum, Oesophagus	Breast Cervix uteri Ovary, oral cavity Colorectum	Breast Lip, oral cavity Cervix uteri Lung Colorectum

quality of diagnosis can be improved for good in upcoming years. The following factors are motivated us to perform this research:

(1) The first factor is the analysis of different types of screening techniques. Various type of screening techniques has a different representation of the targeted area. They have their own pros and cons. This fact motivated us to study the properties of different types of breast cancer screening techniques so that appropriate screening techniques can be selected for breast cancer detection.
(2) The use of deep learning techniques in the identification of breast cancer is another factor. In medical research, deep learning techniques are commonly used. We are studying and analyzing these strategies for breast cancer detection because of the development of novel optimization functions and techniques.
(3) Different types of evaluation parameters are used to validate breast cancer detection techniques. The nature of evaluation measures varies from one to the next. The new technique may perform better on some parameters while performs poorly on others. This feature motivates us to investigate the effectiveness of breast cancer detection evaluation measures.

1.2 Contribution

This paper focuses on the study of breast cancer screening techniques with their pros and cons. This study discusses mainly:

(1) Theoretical aspects of breast cancer for females, males, transgenders including deep learning implementations for the detection of breast cancer.
(2) The different breast cancer screening approaches/techniques, risk factors, target connection, and common datasets.
(3) The mathematical representations of performance evaluation measures
(4) The comparative analysis of deep learning-based breast cancer prediction techniques in terms of performance measures.
(5) The possible future research directions for breast cancer detection

To the best of our knowledge, no review of breast cancer detection involving all of the above mentioned contributions has been reported so far. This study provides a thorough review of the published literature on breast cancer detection screening methods and techniques. Undoubtedly, this study is going to be beneficial for young researchers. Figure 3 depicts the layout of this paper.

The remaining structure of this paper is as follows. Section 2 presents the research methodology used in this study. The datasets for breast cancer detection are discussed in Sect. 3. Section 4 deliberates the performance evaluation measures. Section 5 presents the brest cancer screening techniques. The deep learning techniques for breast cancer detection are mentioned in Sect. 6. Section 7 covers the discussion followed by future research directions in Sect. 8. The concluding remarks are drawn in Sect. 9.

2 Research Methodology

This section presents the survey papers related to breat cancer detection techniques followed by the papere selection and exculsion methodology used in this study.

2.1 Existing Surveys

In the recent past, several research papers were published to summarize the breast cancer detection techniques. The relevant survey papers are discussed as below:

Yassin et al. [25] presented the findings of a systematic review (SR) aimed at determining the current state-of-the-art for computer aided diagnosis and detection (CAD) systems for breast cancer. They provided a broad assessment of CAD systems for image modalities and machine learning-based classifiers. Prospective research studies to develop more objective and efficient CAD systems have been discussed. A brief review of various reported methods and systems for early breast cancer detection was presented by Gupta et al. [26]. A variety of microwave imaging approaches such as microwave tomography and radar-based imaging were investigated. Lu et al. [27] presented some diagnostic imaging methods for breast cancer diagnosis. The breast cancer detection using computer vision and machine learning techniques were investigated. The performance of various methods was analyzed on mammographic images. Huppe et al. [28] presented a comprehensive review on molecular breast imaging. Their research covered the current literature, indications, clinical application, biopsy approach, and MBI integration into medical practice. Oyelade et al. [29] analyzed various deep learning methods for the detection of architectural distortion from digital mammography. The main focus of their study was the detection of abnormalities such as masses and micro-calcification, which are indicators of the disease's advanced stage. Their study indicated that about 70% of the existing literature used Gabor Filters, while only 10% used survey results in computer vision and deep learning to build outstanding computational models for the detection of architectural distortion. Husaini et al. [30] studied the use of thermography and artificial

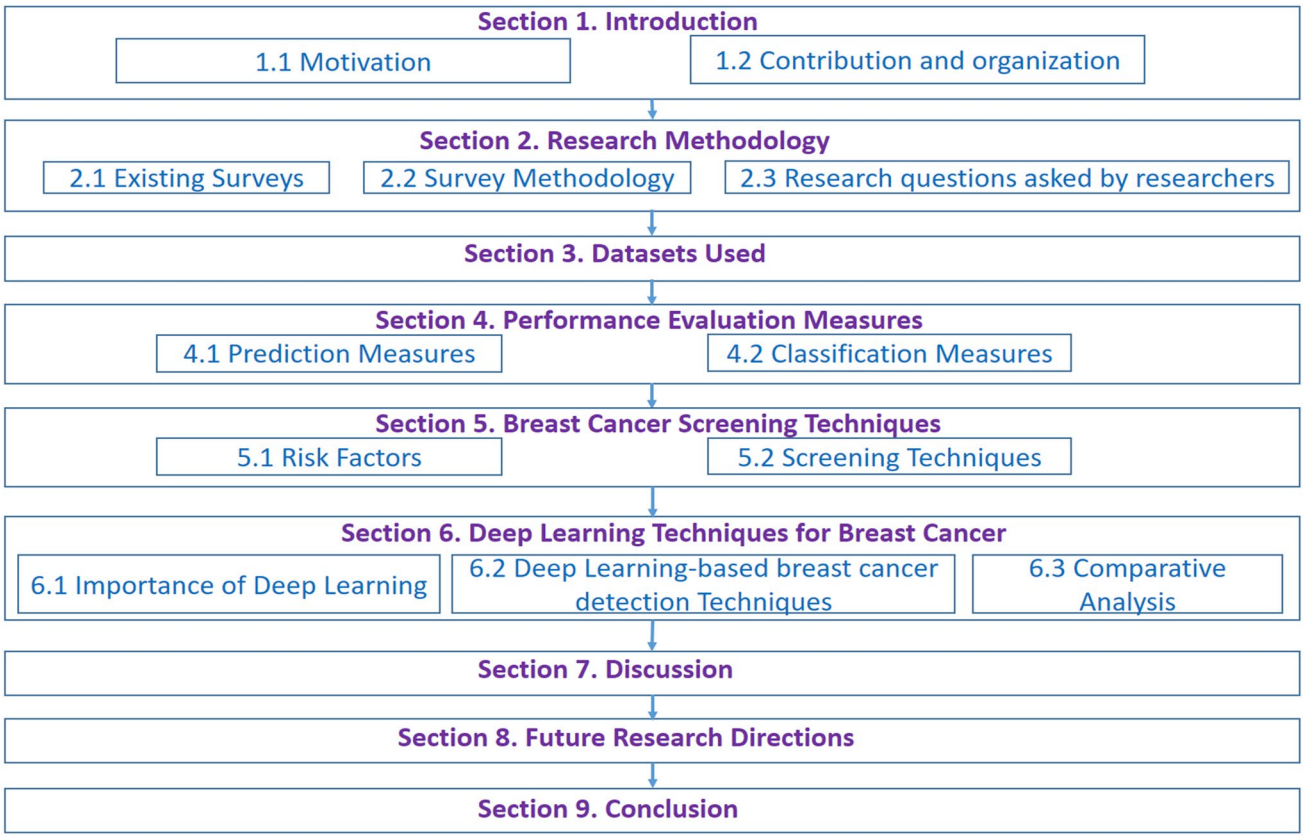


Fig. 3 Structure of paper

intelligence techniques for breast cancer detection. Various deep learning models such as Radial Basis Function Network (RBFN), K-Nearest Neighbors (KNN), Probability Neural Network (PNN), Support Vector Machine (SVM), ResNet50, SeResNet50, V Net, Bayes Net, Convolutional Neural Networks (CNN), Convolutional and De-Convolutional Neural Networks (C-DCNN), VGG-16, Hybrid (ResNet-50 and V-Net), ResNet101, DenseNet and InceptionV3 were analyzed to process thermographic images of breast cancer. Some research works discussed the breast cancer in transgender patients [5, 31]. The qualitative analysis was performed on patient demographics, breast cancer characteristics, breast cancer presentation and management. According to their study, breast cancer present in transgender men is mainly depends upon the top surger.

Due to advancement in deep learning techniques for medical imaging, a need was felt to prepare a survey of research articles summarizing the applications of deep learning techniques for breast cancer detection. Our survey presents the computational studies on breast cancer detection over the last decade, i.e., from 2000 to 2020. The main focus of this study is to investigate the existing breast cancer detection techniques using deep learning, risk factors associated with techniques, and open challenges associated with the existing techniques.

2.2 Survey Methodology

This study on breast cancer detection is conducted through PRISMA [32, 33]. The reason behind the use of PRISMA is that it helps enhance the delineation of CR. It provides the guidance to choose, recognize, and evaluate the studies.

Four different databases namely Google Scholar, Scopus, PubMed, and Preprint plateforms have been used for this study. Four preprint plateforms namely ArXiv, TechRxiv, MedRxiv, and ChemRxiv have been used to conduct the search for appropriate papers. The search string consists of “breast cancer” or “cancer” or “((deep learning) AND (breast cancer))” or “((machine learning) AND (breast cancer))” or “((Artificial Intelligence) AND (breast cancer) AND (detection techniques))” or “breast cancer detection techniques”. Figure 4 shows the search string used for

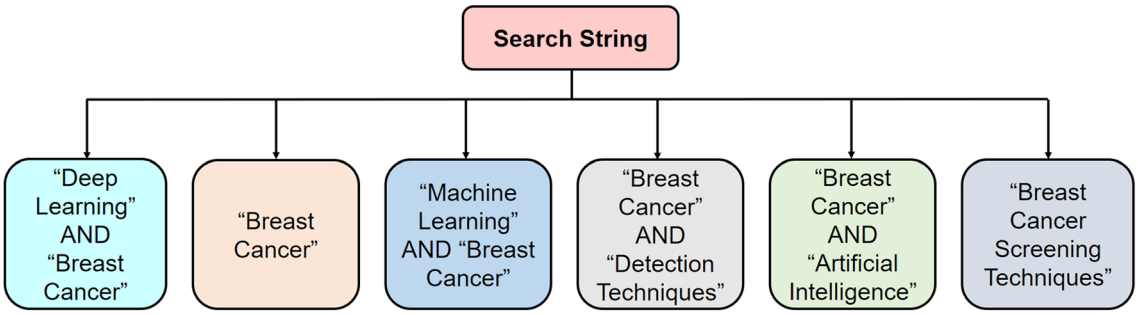


Fig. 4 Search string for searching the research articles

conducting the search. Manual search was also conducted to find out the relevant research papers. In the identification phase, a total of 1600 research publications were chosen.

In the screening phase, 900 research articles were selected after the removal of unsuitable, duplicate, and unrelevant research articles. 750 research articles were excluded after reading the title, abstract, and introduction. The remaining 150 research articles were analyzed through removal criteria and 90 research articles were excluded. Thereafter, 60 research articles were moved to the next phase. In eligibility phase, 30 research articles were eliminated after the evaluation of whole papers. Ultra wideband radar imaging [34], ensemble empirical mode decomposition by ultra-wide band [35], flexible 16 antenna array for microwave [1], Ion-Sensitive Field-Effect Transistor based CMOS integrated Lab-on-Chip system [36] are few related schemes other than deep learning, used for detection of breast cancer. The research articles on these techniques were also eliminated. 30 research articles were designated for review of breast cancer detection techniques. The selection and removal criteria for research articles are mentioned in Table 2. Figure 5 depicts the different phases of PRISMA for this review.

2.3 Research Questions Asked by Researchers

The primary goal of this review is to inspire the young researchers in this area. This paper addresses a number of breast cancer detection-related questionnaire, some of which are listed in Table 3. This will assist innovative researchers in grasping the fundamental concepts of breast cancer detection and determining the open challenges in this field.

Table 4 summarizes the comparsion between the existing surveys and the proposed one in terms of research questions. In this table, denotes that the survey has answered the respective research question, while indicates otherwise.

3 Datasets Used

A variety of datasets is required to develop the computational methods fro breast cancer detection. The datasets are varied in nature. Some datasets have small number of features and tuples. Wheras, some datasets have large number of features and tuples. Researchers use a variety of breast cancer databases for the development and evaluation of computational methods. Some datasets are open to the public and some are limited to specific categories. Table 5 shows the detail description of breast cancer datasets. Table 6 shows the dataset used by researchers in recent years.

Table 2 Selection and removal criteria for selection of research articles

S. no.	Parameter	Selection criteria	Removal criteria
1	Time duration	Research article published from 2010–2021	Research article published before 2010
2	Analysis	Research article including breast cancer detection	Research article including different cancer detection
3	Comparison	Research article focus on deep learning techniques used for breast cancer detection	Research article focus on other techniques used for breast cancer detection
4	Study	Research involving mathematical foundation and experimental results	Research involving case study and articles in different language other than English

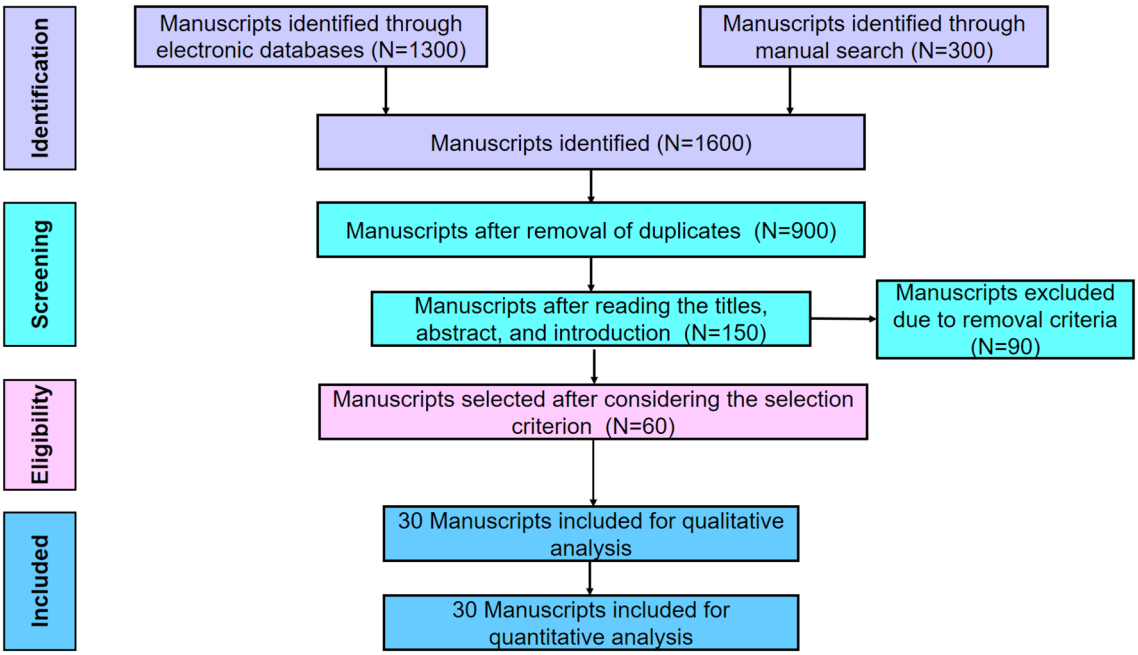


Fig. 5 PRISMA flow diagram on breast cancer review strategy

Table 3 Research questions related to breast cancer detection

Questions	Research questions
Q1	What is breast cancer?
Q2	Explain breast cancer in females, males, trans males, and trans females, along with detection techniques
Q3	What are the risk factors associated with breast cancer?
Q4	What are different types of screening methods involved in breast cancer detection?
Q5	What are different types of deep learning techniques in breast cancer detection?
Q6	What are the performance evaluation measures for validating deep learning based breast cancer detection techniques?
Q7	What are the challenges of breast cancer detection using deep learning?
Q8	What are the future research directions for breast cancer detection using deep learning?
Q9	What is the role of deep learning in breast cancer detection?
Q10	How breast cancer detection using deep learning is different from the other approaches?

Table 4 Comparative analysis of survey papers on breast cancer detection in terms of research questions

Survey	Review year	Research questions									
		Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10
VYassin et al. [25]	2018	✓	✗	✗	✓	✓	✗	✗	✓	✗	✗
Gupta et al. [26]	2020	✓	✗	✗	✓	✗	✗	✗	✓	✗	✗
Lu et al. [27]	2018	✓	✗	✗	✓	✓	✓	✗	✗	✗	✗
Huppe et al. [28]	2017	✓	✗	✗	✓	✗	✗	✗	✗	✗	✗
Oyelade et al. [29]	2020	✓	✗	✗	✗	✓	✓	✓	✓	✗	✗
Husaini et al. [30]	2020	✓	✗	✗	✓	✓	✗	✓	✓		✓
Hartley et al. [31]	2018	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
Stone et al. [5]	2018	✓	✓	✗	✗	✗	✗	✗	✗	✗	✗
Proposed study	2021	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Table 5 Detail description of breast cancer datasets

Datasets	References	Description	Resource	Comment
Mammographic Image Analysis Society (MIAS)	[37]	The database is available on 2.3 GB 8 mm (Exabyte) tape and contains 322 digitized films. The database has been padded/ clipped and reduced to a 200 micron pixel edge, resulting in images that are all 1024 × 1024	Available at the Pilot European Image Processing Archive (PEIPA) at the University of Essex	Improves logistical practice, allows for the use of computer-aided detection programs, and has a cancer detection rate comparable to that of the screen- lm mammography. It's unclear how this will affect recall rates
Digital Database for Screening Mammography (DDSM)	[38]	The database is now fully functional, with 2620 cases. This is a mix of benign without callback, benign, normal and cancer volumes that were specially selected and digitized for DDSM	The research com- munity has made extensive use of the DDSM. It is kept at the University of South Florida so that it can be accessed via the inter- net	
IN breast	[39]	The images were taken at the Breast Centre in CHSJ, Porto, between 2008 and 2010, with approval from both the Hospital's Ethics Committee and the National Committee for Data Protection	IN Breast dataset can be requested online at http://medicalresearch.inescporto.pt/breastresearch/index.php/GetINbreast Database	
US1	[40]	The data came from an expert didactic media le for breast screening specialists in 2001. The dataset contains 306 images with a mean image size of 377 × 396 pixels from various cases	To acquire this dataset, the user must purchase it from Prapavesis et al	Requires a skilled operator, the examination technique is not standardized, the interpretation criteria are variable, and micro calcifications are not detected
US2	[41]	This data was gathered in 2012 from the Parc Taul' Corporation's UDIAT Diagnostic Centre in Sabadell (Spain). The database contains 163 images from various women, with an average image size of 760 × 570 pixels	The breast lesions dataset is available on the internet (goo.gl/SJmoti) for research purposes	
Breast Ultrasound Dataset	[42]	Breast ultrasound images from women aged 25 to 75 years old were collected at the start of the study. This information was gathered at Baheya Hospital in 2018. The total number of female patients is 600. There are 780 images in the database, with an ordinary image size of 500 × 500 pixels	Publically Available at https://cholar.cu.edu.eg/?q=afahmy/page s/dataset [43]	

Table 5 (continued)

Datasets	References	Description	Resource	Comment
Reference Image Database to Evaluate Therapy Response (RIDER)	[44]	Data was gathered in order to reach an initial agreement on how to harmonize collection of data and analyze quantitative imaging techniques for assessing drug or radiation therapy response	Publically available at https://wiki.cancerimageingarchive.net/display/Public/RIDER+Breast+MRI	In women at increased risk, it is more sensitive and marginally less specific than mammography. No radiation necessitates intravenous contrast, which is time-consuming and inconvenient for some women, such as those who have a pacemaker, aneurysm clips, or claustrophobia
QIN Breast DCE-MRI	[45]	The breast DCE-MRI data set used for study was collected provisionally under a HIPAA (Health Insurance Portability and Accountability Act of 1996)-compliant, Approval From the institutional Board with the exemption of consent	Publically available at https://wiki.cancer i magingarchive.net/display/Public/QIN+Breast+DCE-MRI	
DBT-TU-JU	[46]	There are 1100 Breast thermograms of 100 subjects in the DBT-TU-JU database. This research reflects the generation of ground truth images of the hotspot areas, whose presence in a breast thermograms indicates the presence of breast abnormality, due to the necessity of evaluating any breast abnormality detection system	Only patients and their physicians have access to private databases that are only accessible for internal purposes	Not harmful, non-invasive, time consuming, requires more work to be done, still in practice
DMR: Database For Mastology Research	[47]	Thermal and mammography images obtained by our research group are among the images available in this database. For a sample of 64 breasts, the accuracy was 100% (composed of 32 healthy and 32 unhealthy)	The images are stored in the Database for Research Mastology with Infrared Image-DMR-IR, accessible on the website http://visual.ic.u.br/dmi	

Table 6 Breast cancer datasets used by researchers in recent years

References	Year	Dataset									
		MIAS	DDSM	INbreast	US1	US2	BUSI	ABUS	DCE-MRI	DBT-TU-JU	DMR
[48]	2015	✗	✗	✗	✗	✗	✗	✗	✓	✗	✗
[49]	2016		✗	✗	✗	✗	✗	✗	✗	✗	✗
[47]	2017	✗	✗	✗	✗	✗	✓	✗	✗	✗	✗
[50]	2018	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[41]	2018	✗	✗	✗	✓	✓	✗	✗	✗	✗	✗
[51]	2019	✗	✗	✗	✗	✗	✗	✗	✗	✗	✓
[52]	2019	✗	✗	✗	✗	✗	✗	✗	✗	✓	✗
[53]	2020	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗
[54]	2020	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
[55]	2020	✗	✓	✓	✗	✗	✗	✗	✗	✗	✗

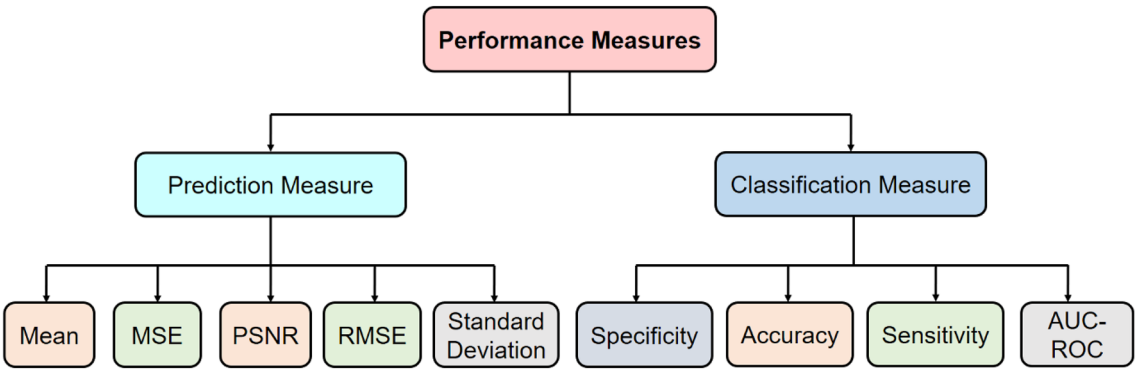


Fig. 6 Classification of performance measures

4 Performance Evaluation Measures

Different performance measures are used to evaluate the performance of breast cancer prediction models. The performance measures are generally claasified into two main categories namely, prediction and classification measures [56]. Figure 6 shows the classification of performance measures.

4.1 Prediction Measures

Mean, standard deviation, mean square error (MSE), root mean square error (RMSE), and peak signal to noise ratio (PSNR) are the well-known prediction measures. The mathematical formulation of these measures is given in succeeding subsections.

4.1.1 Mean (μ)

Mean represents the average brightness of an image. If the average intensity of a breast cancer image is much high, then the density of tissue is also high. The mathematical formulation of mean (μ) is defined as [57]:

$$\mu = \frac{1}{mn}\sum_{a=1}^{m}\sum_{b=1}^{n} pA(a,b), \tag{1}$$

where m and n signify the number of rows and columns in an image. pA is the coefficient of approximation. The value of mean should be high for better results [57].

4.1.2 Standard Deviation (σ)

Standard deviation (σ) can be described as a measure of how much the contrast intensity increases when the texture irregularity increases [58]. It is defined as:

$$\sigma = \sqrt{\frac{1}{mn}\sum_{a=1}^{m}\sum_{b=1}^{n}(pA(a,b)-\mu)^2} \tag{2}$$

4.1.3 Mean Square Error

The differential between observed and predicted values is measured using the mean square error (*MSE*) [59]. The mathematical formulation of *MSE* is given below:

$$MSE = \frac{1}{mn}\sum_{a=1}^{m}\sum_{b=1}^{n}(I_o - I_p)^2, \tag{3}$$

where I_o and I_p denote the observed and predicted values, respectively.

4.1.4 Root Mean Square Error

The square root of second moment of difference between the observed and predicted values is known as root mean square error (*RMSE*) [60]. It can be defined as the standard deviation of prediction errors.

$$RMSE = \sqrt{\frac{1}{mn}\sum_{a=1}^{m}\sum_{b=1}^{n}(I_o - I_p)^2} \tag{4}$$

RMSE is a reliable indicator of the accuracy obtained from the prediction model. *RMSE* has a non-negative value at all times. It is proportional to the scale. It is sensitive towards the outliers.

4.1.5 Peak Signal to Noise Ratio

Peak signal-to-noise ratio (PSNR) is the ratio of an image's maximum achievable power to the power of degrading noise that influences its representation quality [61].

$$PSNR = 10\log_{10}\frac{(l-1)^2}{MSE}, \tag{5}$$

where l is the number of highest allowable intensity levels in an image.

4.2 Classification Measures

The performance of breast cancer prediction model is evaluated by using classification measures. These measures are positive predictive value, sensitivity, accuracy, specificity, and area under receiver operating characteristics. The mathematical formulation of these measures is mentioned in the succeeding subsections.

4.2.1 Positive Predictive Value (PPV)

Positive predictive value (PPV) is the fraction of suitable instances among the recovered instances [62]. It is also known as Precision and is defined as [62]:

$$PPV = \frac{TP}{TP+FP}, \tag{6}$$

where TP is the number of true positives and FP is the number of false positives, respectively. The true positives are the positive tuples that the prediction model accurately predicts. The false positives are the negative tuples that the model predicts incorrectly. The value of PPV lies in the range of [0, 1].

4.2.2 Sensitivity

Sensitivity (S_n) is a metric for assessing the efficacy of breast cancer detection prediction models. S_n is also known as the rate of recognition [62]. It specifies the percentage of positive tuples that the prediction model successfully predicts.

$$S_n = \frac{TP}{TP+FN}, \tag{7}$$

where FN shows the number of false negatives. The false negatives are the positive tuples that the prediction model predicts incorrectly.

4.2.3 Accuracy

The percentage difference of projected synergy scores from observed results within the allowable error range is called accuracy [62]. It is defined as:

$$A_c = \frac{(TP+TN)}{(TP+TN+FN+FP)} \times 100, \tag{8}$$

where TN represents the number of true negatives. The term "true negative" refers to negative tuples that the prediction model accurately predicts.

4.2.4 Specificity

True negative rate is used to describe the specificity. It refers to the percentage of negative tuples properly predicted by the prediction model [63].

$$S_f = \frac{TN}{TN+FP} \tag{9}$$

4.2.5 Area Under Receiver Operating Characteristics Curve

Receiver Operating Characteristic (ROC) represents the trade-off between true positive rate (TP_r) and false positive rate (FP_r) [63]. The false positive rate and true positive rate are represented by the x-axis and y-axis of ROC curve, respectively. The area under ROC curve (AUC-ROC) is a metric for computing the model accuracy. The value of AUC-ROC lies in the ranges of [0.5, 1].

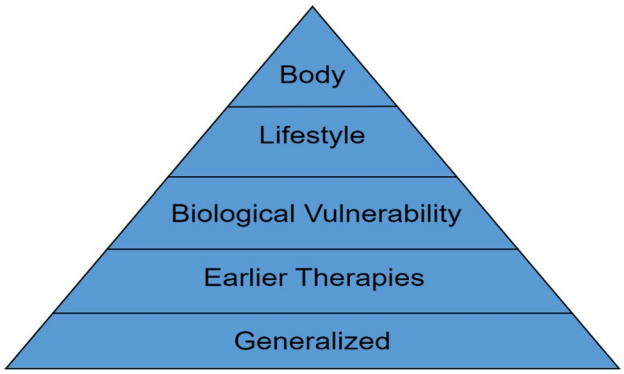


Fig. 7 Risk factors associated with breast cancer

$$TP_r = \frac{TP}{TP + FN} \tag{10}$$

$$FP_r = \frac{FP}{FP + TN} \tag{11}$$

5 Breast Cancer Screening Techniques

This section discusses the risk factor associated with breast cancer followed by the breast cancer screening techniques.

5.1 Risk Factor Associated with Breast Cancer

According to the American Cancer Society, many factors are responsible for enhancing the likelihood of breast cancer [7]. Figure 7 depicts the risk factor associated with breast cancer. The well-known risk factors are age, family history, reproductive factors, earlier therapies, and lifestyle. The detail description of these factors is mentioned in Table 7.

5.2 Screening Techniques

Breast cancer awareness provides help to the affected people so that they can take better decisions about their health.

Table 7 Description of risk factors related to breast cancer

Risk factor	Factor	Explanation
Generalized	Gender	Being a female is perhaps the most critical risk factors for breast cancer growth
	Age	The risk of breast cancer rises as the woman grows older
	Race	Pretty much across the globe, white women tend to get breast cancer marginally more often than African-American women
Body	Menstrual history	The risk of breast cancer is marginally higher among women who start menstruation early (before age 12) and/or menopause early (after the age 55). This rise in risk may have been caused by the progesterone and estrogen hormones being released longer in life
	High breast density	Dense breast tissue means more tissue and less tissue is contained in the gland. The risk of breast cancer is higher for women with denser breast tissue
	Not having offspring	Females who do not have babies or who were later pregnant might be more likely to develop breast cancer. Breastfeeding may contribute to reducing your risk of breast cancer
	Weight	Fat tissue can increase estrogen after menopause, and high estrogen levels can increase the risk of breast cancer. Adult weight gain and excess corporeal fat may also be significant around the waist
Lifestyle	Inactive Lifestyle	Breast cancer risk reduction is helped by physical activity
	Alcohol	Increased risk of breast cancer is associated with consumption of alcohol. With alcohol consumed, the risk increases
Earlier therapies	Therapy with DES	The risk of breast cancer is marginally higher for women who have been given DES (diethylstilbestrol) in the course of pregnancy
	Hormone treatment after menopause	The risk of breast cancer is raised by the use of estrogen and progesterone during menopause
Biological vulnerability	Family Background	A mother, sister or daughter who experiences breast cancer may increase the risk
	Ancestral Factor	Hereditary modifications (genetic changes) may increase risk in certain genes, such as BReast CAncer gene (BRCA) 1 and 2

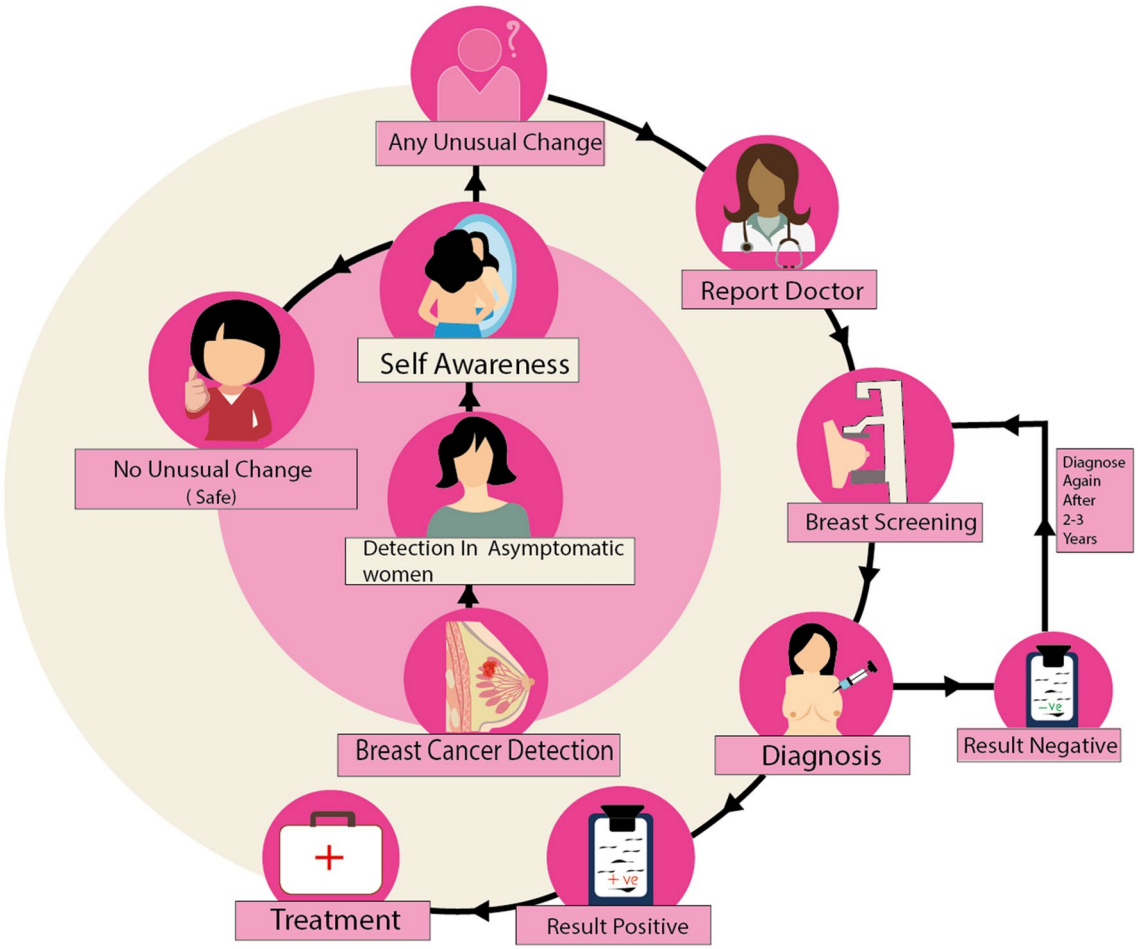


Fig. 8 Breast Cancer detection for asymptomatic women

Hormonal changes, genetics, breast density, and lifestyle are greatly affected the appearance of breasts. Understanding the morphology and physiology of usual breast tissue is an essential for predicting the any development of breast cancer and may aid in the early diagnosis of abnormal lesions. Figure 8 shows breast cancer detection in asymptomatic females.

Breast cancer diagnosis entails a variety of screening techniques to improve the accuracy of diagnosis. The well-known breast cancer screening techniques are X-ray mammography, breast ultrasound, Magnetic Resonance Imaging (MRI), and Positron Emission Mammography (PEM) [8]. X-ray mammography is the most effective technique. Breast Ultrasound uses sound waves to create a picture of tissues inside the breast. However, this technique suffers from low specificity, high cost, and a lack of availability [23]. The accuracy of breast ultrasound is approximately 67.8%. MRI is the another technique for breast cancer detection. It creates detailed images of organs by combining a large magnet, radio waves, and a computer. This technique has a higher sensitivity, however, high cost and low specificity that can lead to overdiagnosis [23, 64, 65]. The accuracy obtained from MRI may lie in the range of 70% to 72%. PEM is an alternate method for breast cancer screening [66]. It has high specificity as compared to the other techniques. However, it suffers from low sensitivity and high radiation expose. Breast cancer screening techniques are broadly categorized into three main groups such as physical, electrical, and mechanical (see Fig. 9).

5.2.1 Physcial Screening Techniques

The well-known physical screening techniques are mammography, ultrasound, and MRI. The detail description of these techniques are given in the succeeding subsections.

5.2.1.1 Mammography Breast self-examination (BSE), Clinical breast examination (CBE), and mammography are commonly used screening techniques for breast cancer detection [67, 68]. Nowadays, digital mammograms are widely used for breast cancer detection. In digital mammogram, X-rays are replaced with solid-state detectors that

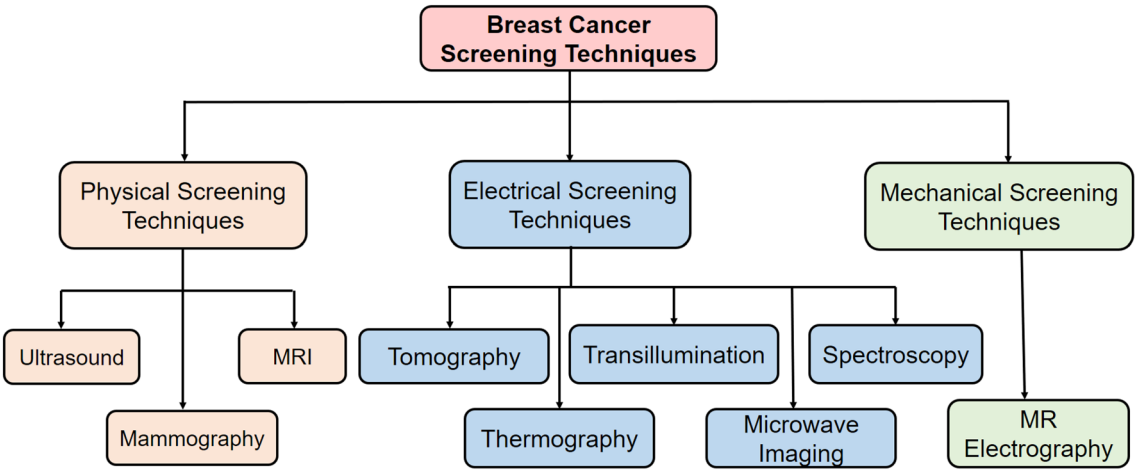


Fig. 9 Classification of breast screening techniques

translate X-rays into electrical signals. It is also known as full-field digital mammography (FFDM). The detectors in digital cameras are identical. Electrical impulses are used to create breast images and displayed on a computer screen [69]. Computer-aided detection (CAD) solutions are created to read mammographs. CAD systems usually interpret a mammogram and identify questionable places, which are investigated by the radiologist [70].

Ribli et al. [50] proposed a CAD program focused on Faster R-CNN. This program was able to classifiy malignant or benign tumors in a mammogram without user intervention. Wang et al. [53] proposed an end-to-end method for mammographic diagnosis. This approach eliminated the manual preprocessing. The treatment of mammograms was introduced in one situation with a different approach focused on Multiscale (MS) system and Multi-Instance (MI) system. MS module selects the basic features of mammograms and MI module took the general situation into account in one event. The output of these modules are combined to get the better results. Heidari et al. [71] introduced a new computer-aided diagnosis (CADx) scheme based on analysis of global mammographic image features. This research demonstrates the possibility to build a modern CADx mammogram high-performance global picture processing scheme. This technique is more effective and reliable than the previous techniques. Ekici et al. [51] utilized a convolution neural network (CNN) for thermographic breast cancer screening. They used five different processes namely, data acquisition, image processing, segmentation, extraction, and classification. CNN provides better results than the other techniques in terms of prediction results.

5.2.1.2 Ultrasound Breast ultrasound is a common way to test for breast cancer as it allows the screening sensitivity can be increased in thick breasts [72–74]. Wang et al. [54] developed an automatic breast cancer diagnosis of Automated Breast Ultrasound System (ABUS). They showed a lack of control for the existence of the adaptive barrier at voxel stages for cancer and non-cancer patients. The suggested network allows for an effective breast cancer screening scheme by utilizing ABUS with small false positives. Shen et al. [75] studied the comparative analysis of both mammography and breast ultrasound. It is observed that China's breasts appear to be tiny and compact. Ultrasound is a standard tool for screening the breast cancer in China. For clinical experiments, though, the efficacy and risks of mammograms were not measured. Nyayapathi et al. [76] introduced a new photoacoustic tomography device that displays angiographic features with mammogram-like images in the breast. The mechanism portrays a highly compact breast of two flat, 2.25 MHz transceiver clusters of 128 components, and line optical fiber bundles from top to bottom. The soft compression is done using silicone prints, which allows the woman more relaxed than hard metal plate used with conventional mammograms. Dual Scan Mammoscope (DSM) technology developed in this study acquired both ultrasound and photoacoustic breasts.

5.2.1.3 Magnetic Resonance Imaging Magnetic resonance imaging (MRI) plays a significant role in medical field. The scan in this method is used to produce detailed images of the inner body utilizing intense magnetic fields and radio waves [77]. MRI scans are being used to investigate nearly any area of the body including the brain, bones, breasts, heart, and even internal organs. MRI is widely used in breast cancer imaging. Sun et al. [78] identified glioma classification methods for the prediction of radiomics feature. MRI extracted quantitative features from tumor areas. The modality of extraction by radiomics was greater than the other combinations of tumor area.

Whereas, Li et al. [35] used ultra-wideband microwave imaging for early breast cancer detection. They also used an ensemble empirical mode decomposition (EEMD) for direct extraction of tumor. Only isolated signals from as-detected waveforms are required for the reconstruction of the picture for tumor detection. They used MRI to create more precise models for electromagnetic analysis. Here, a tumor of 4 mm in diameter within the glandular or at the interface between fat and the gland has been shown by the proposed procedure. In case the glandular tissue has a bigger dielectric constant of 35, tumor reaction may also be identified. Their research showed that the solution presented could serve as an important alternative to direct tumor response extraction. Mahrooghy et al. [48] suggested the spatiotemporal dynamic properties of wavelet from dynamic contrast-enhanced magnetic resonance imaging (DCE-MRI) to quantify breast cancer intra-tumor heterogeneity.

The receiver operating characteristic (ROC) and area under the curve (AUC) were computed to assess the performance of classifier using leave-one-out (LOO) cross-validation. The heatwave features outperformed the other features. The combination of heatwave and standard features can further enhance the performance of classifier [62]. Table 8 summarize the physical scrrening techniques for breast cancer.

5.2.2 Electrical Screening Techniques

Impedance spectroscopy, thermography, transillumination, microwave imaging, and tomography are the well-known electrical screening techniques. The detail description of these techniques are given in the succeeding subsections.

5.2.2.1 Impedance Spectroscopy Researchers are using this technique to improve breast cancer detection techniques. Haeri et al. [82] introduced two experimental breast cancer screening instruments namely, electrical impedance spectroscopy (EIS)-Probe and the EIS-Hand-Breast (EIS-HB). EIS-Probe and EISHB system were able to assess the electrical properties of breast tissue. Cancerous tissues were identified by determining the change in parameters of healthy tissues. Huerta-Nuñez et al. [83] utilized the bioimpedance spectroscopy to investigate the cancer cells in an aqueous media. Experimental results revealed that impedance spectroscopy has a sufficient sensitivity for identification of extraordinarily low cancer cell composition in an aqueous solution. Lederman et al. [84] designed a seven-probe resonance-frequency-based electrical impedance spectroscopy (REIS) system and used the data of 174 females. Artificial neural network (ANN), support vector machine (SVM), and Gaussian mixture model (GMM) were used. The results revealed that ANN attained the maximum values of ROC and AUC as compared to the other classifiers. REIS examinations provide the relevant information to build classifier for the stratification of breast cancer risk. Ward et al. [85] evaluated the inter-arm impedance ratio range for evaluating the value of threshold as a standard for detecting lymphedema associated with breast cancer. When an impedance of 1106 is surpassed by a danger to the neuronal limb and 1134 when the dominant limb is at danger, relative to those currently in use of 1066 and 1139, the existence of lymphedema is recorded. The variation in these values can be considered as the minor significance towards the clinical practice.

5.2.2.2 Thermography A special camera is used to measure skin temperature on the surface of the breast and is known as thermography. It is a non-invasive and radiation-free research [86]. Ekici et al. [51] developed an automatic breast cancer detection technique. They used image processing and analytics techniques to analyze the thermal images of breast. The feature extraction algorithm was proposed to extract the features for identification of breast images as regular or suspicious. Their technique attained the accuracy of 98.95% for the thermal images of 140 females. Jose-Luis et al. [87] proposed a technique to solve the inverse thermal transfer problem in the Levenberg Marquardt algorithm. This technique was used to identify and locate malignant tumors within the breast using a patient-specific digital breast model and clinical infrared imaging (IRI) images. Digital heat amplification systems were used to tackle the challenges occurred during the identification of size and position of malignant tumor within the breast. This technique can be combined with mammography to detect the breast cancer, especially in the case of dense breasts. In [52], the advancements in thermography-based techniques were investigated for breast cancer detection. It is observed from breast thermograms that breast cancer signs can be detected through the asymmetrical thermal spreads between breasts. Their study showed that the neural network systems enhanced the prediction accuracy of breast cancer thermograms. Mambou et al. [88] explored an infrared digital imaging techniques for breast cancer. The basic assumption in this technique is that the increase in thermal activity in the precancerous tissues and the areas surrounding developing breast cancer. They concluded that infrared image processing techniques require a CAD system for detection. Roslida et al. [76] studied the three convolutional neural network (CNN) models namely, ResNet101, DenseNet201, MobileNetV2, and ShuffleNetV2 for breast cancer detection. Database for Mastology Research (DMR) was used to evaluate the performance of above-mentioned models. DenseNet201 was capable to classify both static or dynamic images. Transillumination.

Transillumination is a procedure used in an organ or part of the body to detect anomalies. The examination is

Table 8 Classification of breast cancer detection using physical screening methods

References	Classification model	Performance measures	Challenges	Advantages
[50]	Pre-processing + Faster R-CNN + Regional proposal network	AUC = 0.85 lesion detection sensitivity = 90%	The detection performance could only be evaluated on the small INbreast dataset	Model reaches high sensitivity with very few false positive marks
[53]	Pre-processing and image cropping + CNN for feature selection + Multi-instance module + final diagnosis of the whole case	AUC = 0.865	In future author could assist other advanced applications, such as size measurement, lesion characterization	Their method learned the unique features of lesions via the multiscale module
[49]	Pre-processing + CS-LBP features from each 2 × 2 blocks in Wavelet Domain + SVM-RFE based feature selection + Random Forest classification	Accuracy: 97.25%, Precision: 97.3%, Recall: 97.2%, F-measure: 97.2% MCC: 94.1% ROC Area: 97.61%	The author can further classify the type of tumor	Fast feature extraction, small feature dimension, fast modelling and prediction
[55]	Image enhancement and segmentation + DenseNet169 for feature extraction + Region based Group-max Pooling + Prediction	(INbreast)RGP: AUC = 0.934 and GGP: AUC = 0.924 (CBIS-DDSM) RGP: AUC = 0.838 and GGP:AUC = 0.823	The results of visualization show that the proposed model can roughly locate suspicious regions	Ability of learning lesion location information
[79]	Pre-processing + feature extraction + deep multi-view classifier	AUC = 0.895	AUC could be increased	Their model could improve radiologist sensitivity for breast cancer detection
[54]	Pre-processing + pre-trained UNet + DDS feature extraction + TM	Sensitivity = 95% with 0.84 false positives per volume	In future author could assist other advanced applications, such as size measurement, lesion characterization	High sensitivity and low false positives
[71]	Image segmentation + SSIM feature extraction + DCT feature extraction + FFT feature extraction + SVM classifier	AUC = 0.85–0.91(from one of the three sub categories) AUC = 0.96 ± 0.01(in three sub categories)	Its clinical utility or impact on radiologists' performance in diagnosis of breast cancer using mammograms has not been tested	significantly higher performance with AUC
[9]	HA-BiRNNs	AUC: 0.8854 REC: 0.8771 F1 score of prediction:0.9070	The author wish to develop methods to interpret the representation learned by Knowledge-powered Deep Breast Tumor Classification model to classify malignant and benign tumor pattern	Method achieves higher performance
[35]	Pre-processing + canny edge detection technique for boundary selection + antenna arrangement and simulation process + extraction of tumor response signals + result	N/A	Their method offers the efficient detection even for a tumor of 4 mm diameter located within the glandular or at the interface between the gland and fat	They state that their approach can be an effective alternative to direct extraction of the tumor response
[48]	Pre-processing + DCE-MRI (wavelet transform) + heatwave feature extraction + classification	(AUC = 0.88 HetWave vs. 0.70 standard features). The combination of HetWave and standard features further increase classifier performance (AUCs 0.94)	HetWave could assist other advanced applications such as feature extraction approach for characterizing tumor heterogeneity, providing valuable prognostic information	Superior ROC AUC
[80]	Data acquisition + image preprocessing + co-occurrence analysis computing framework + neural network classifier + hypothesis testing	TPF = 0.90 FPF = 0.09	Author could help with more advanced applications in the future, such as size measurement and lesion characterization	It investigate heterogeneous tumors by separately distinguishing the benign and necrotic tissues within a lesion having malignancy

Table 8 (continued)

References	Classification model	Performance measures	Challenges	Advantages
[81]	Pre-processing + fuzzy c-means thresholding based image segmentation + hybrid feature extraction + feature selection + classification	SVM-MLP classifier: 87% accuracy, 95% sensitivity, and 75% specificity SFS-KNN: 87.50% accuracy, 95.83% sensitivity, and 62.50% specificity	Some SVM classifier's are not capable of classifying the negative sample	GA-MI based feature selection the SVM classifiers for MLP, linear, and quadratic classifier are performing better
[23]	Pre-processing + construction of biomechanical model + image registration + image fusion + analysis of USCT images + result	–	The current limitations of their method is that overall accuracy is not improved with heterogeneous stiffness distribution models	Smaller training and test errors rates

conducted in a dark room with a light-reflecting on a particular body segment to look under the skin [89]. It is an invasive method and is not being used much nowadays.

5.2.2.3 Microwave Imaging Microwave imaging is a promising method for detecting the early-stage breast cancer [90]. Li et al. [35] proposed a direct tumor response extraction technique based on the ensemble empirical mode decomposition for early breast cancer detection. The extracted signals were used to reconstruct the image for tumor detection. diFlorio [91] designed some enhancements in both hardware and software for microwave breast imaging. The hardware monitors the signals down to sub-centimeter screen resolution compatible with a test time of fewer than 2 min. The software resolves the huge time workload and produces accurate images in less than 20 min. They were able to produce the first microwave tomographic images. Klemm et al. [90] studied the imaging of inhomogeneous breast phantoms for microwave breast cancer. They introduced an image enhancement algorithm, which utilizes the concepts of the delay and sum algorithm (DAS) and coherence factor. Their proposed approach was able to reduce clutter and provide better images as compared to the previous techniques. Tuncay et al. [92] presented an effective way to design 3D microwave models. Yin et al. [34] suggested ultrawideband radar imaging for the breast cancer detection. Robust and Artifact Resistant (RAR) algorithm was developed to overcome the negative effects of both artifact and glandular tissues. RAR enhanced the identification capacity, robust artefact resistance, and high detection range.

5.2.2.4 Tomography Tomography is a technique that creates images of single planes of tissue. Kao et al. [94] studied Electrical Impedance Spectroscopy (EIS) to locate and differentiate cancer from normal tissues and benign tumors. The tumors are different from the normal tissue in terms of their conductivity and permittivity. The high contrast tissue, occurs between several kHz and several MHz, can be able to distinguish the malignant from benign. In a silicone phantom breast, the system can detect a 10 mm tumor in a silicone phantom breast. Baran et al. [95] investigated the potential clinical usability phase-contrast micro-computed tomography (micro-CT) with high spatial Resolutions. SYRMEP beamline of the Elettra Synchrotron was scanned with 10 breast tissue specimens of 2 mm in diameter using the phase-contrast micro-tomography propagation method. The high-resolution images was able to provide the detail tissue design assessment at a close-to-histological level. Table 9 summarizes the electrical screening techniques in terms of performance measures and datasets.

Table 9 Classification of breast cancer detection using electrical screening techniques

References	Classification model	Performance measures	Challenges	Advantages
[93]	Image-Based 3-D Surface Reconstruction + Model-Based Segmentation of Breast + Data Collection and accuracy evaluation + Result	The system can detect a 10 mm tumor in a silicone phantom breast	A limitation of these dense optical flow techniques is their sensitivity to lighting variation	Their system can reconstruct the breast surface with average errors of less than 1 mm
[82]	Two innovative instruments' setups for early detection of breast cancer were used	Method expose promising validity in comparison to other breast cancer detection tools	The LSM error function has not been used due to its higher sensitivity	Promising validity
[51]	Data acquisition + preprocessing + segmentation + feature extraction + CNN classifier	98.95% Testing accuracy	The author can further classify the type and size of tumor	Obtained good accuracy rate was obtained for the thermal images
[87]	Breast imaging + Patient specific digital breast model + extract temperature of region of interest + levenverg-marquardt algorithm	The screening technique is found to be more accurate and harmful radiation free	Requires more computation time	The technique has potential to be an accurate adjunct to mammography
[91]	combination of hardware and software to create 3-D microwave tomographic images of the breast	The hardware used in this work has the capability of gathering data up to 3 GHz	The results can be improved by operating at higher frequencies and/or using a multi-frequency approach over an ultra-wide band	Its advantage is its specificity driven by the wide range of dielectric properties
[92]	Preprocessing of MRI data + Bias Field Correction + Segmentation of Two Main Tissues + Electromagnetic Properties Mapping + Building the 3-D Structure	The test conducted was successful	Work could be done to improve correlation between mammographic density and the MRI density	More consistent to classify ARN-MBPs with BTI score rather than ACR classification
[84]	Breast REIS + mirror matched feature extraction + ANN/SVM/GMM + fusion	ANN classifier was found to be the best single classifier among the three tested classifiers, with an AUC of 0.81 and sensitivity of 75% at 80% specificity	The main limitation of their work is smaller sample size	The REIS-based classification decisions are consistent with biopsy recommendations

Table 10 Classification of breast cancer detection using MS Elastography

References	Classification model	Performance measures	Challenges	Advantages
[96]	Acoustic shear waves generating device + elasticity imaging + an algorithm for processing the wave images to generate quantitative images depicting tissue stiffness + prototypic breast MR elastography technique + results	The results confirm the hypothesis that the prototypic breast MR elastographic technique can quantitatively depict the elastic properties of breast tissues in vivo and reveal high shear elasticity in known breast tumors	Considerable scope exists for technical improvement to determine the possible performance of an optimized MR elastographic technique in terms of resolution and quantitative accuracy	Their work shows that it is feasible to use a technique combining MR imaging and acoustic technologies

5.2.3 Mechanical Screening Techniques

MR elastrography is a well-known mechanical screening technique. The detail description of this technique is mentioned in the succeeding subsection.

5.2.3.1 MR Elastrography In [96], electromechanical operators vibrate the breast in MR electrography and produce the acoustic sound waves. An algorithm was used to produce the quantative images from these waves. This technique was evaluated on six healthy persons and six patients. then described by MRI [96]. Goddi et al. [97] presented a review on breast elastography. They discussed future techniques, which are not yet in clinical practice. Table 10 depicts the classification of breast cancer detection using MS Elastography.

6 Deep Learning Techniques for Breast Cancer

This section presents the importance of deep learning in the field of breast cancer followed by the classification of deep learning-based breast cancer detection techniques.

6.1 Importance of Deep Learning in Breast Cancer

The literature reports that machine learning is widely used technique for breast cancer research. K-Nearest Neighbor (KNN), support vector machines (SVM), and naive bayes classifier perform better in their respective fields. However, the tracking and detection processes in machine learning are done manually. For efficient cancer detection, the system needs to process 200 to 300 cells per frame, which is not possible through manual tracking. Hence, there is a need to develop the efficient methods for breast cancer detection. Whereas, deep learning can identify the complex patterns in raw data. Nowadays, deep learning is widely used to identify the breast cancer. According to a study published in Nature Medicine, deep learning models are capable of detecting breast cancer 1 to 2 years earlier than those with the standard clinical methods [98] would have. Deep learning models can learn the most relevant features to solve the problem optimally. Due to this, deep learning models can serve as the best hierarchical feature extractors [99]. The above-mentioned facts motivate the researchers to use learn and hence apply the deep learning techniques for breast cancer detection.

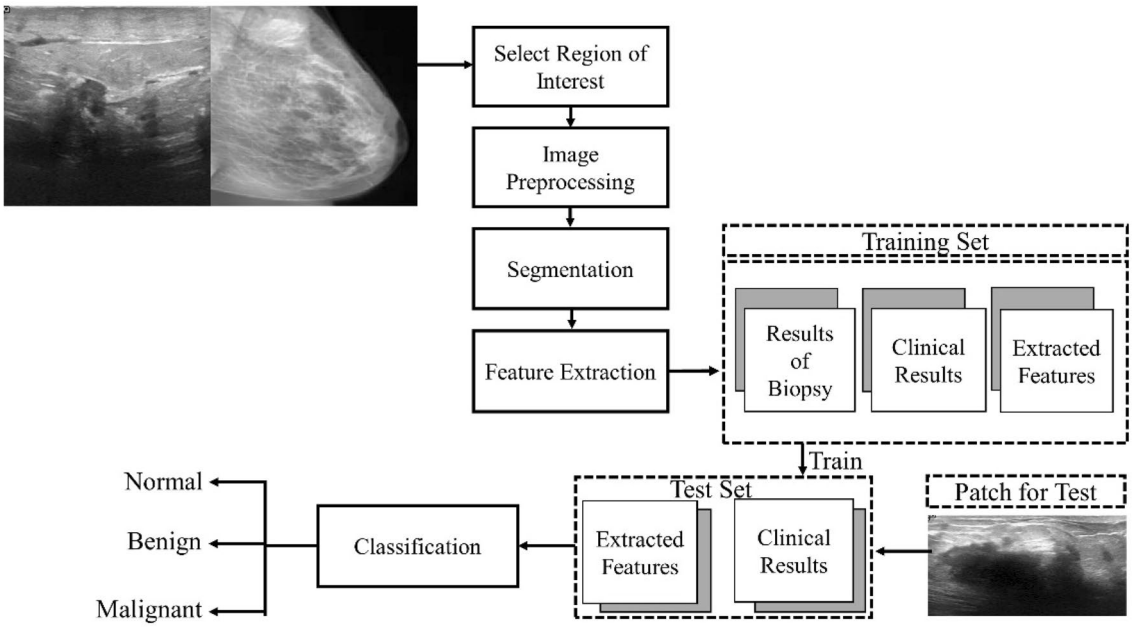


Fig. 10 General framework for breast cancer detection using deep learning

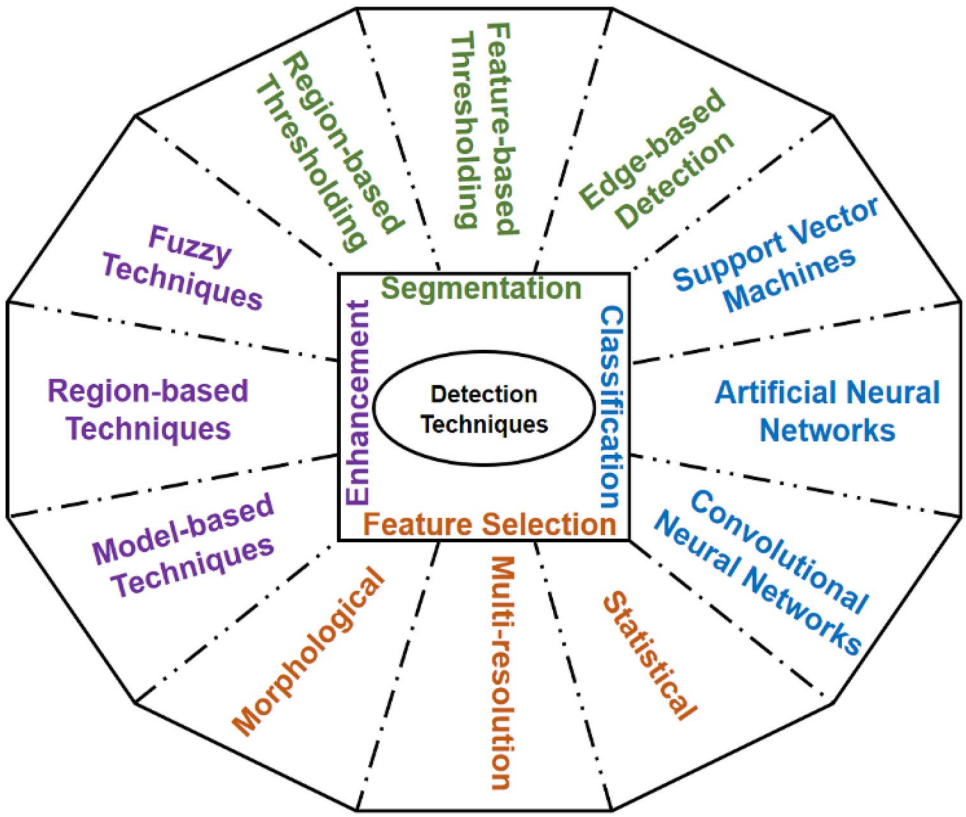


Fig. 11 Classification of Breast Cancer Detection Techniques

6.2 Deep Learning-Based Breast Cancer Detection Techniques

Deep learning architectures are successfully used in the detection of breast cancer. Figure 10 shows the general framework for breast cancer detection using deep learning technique.

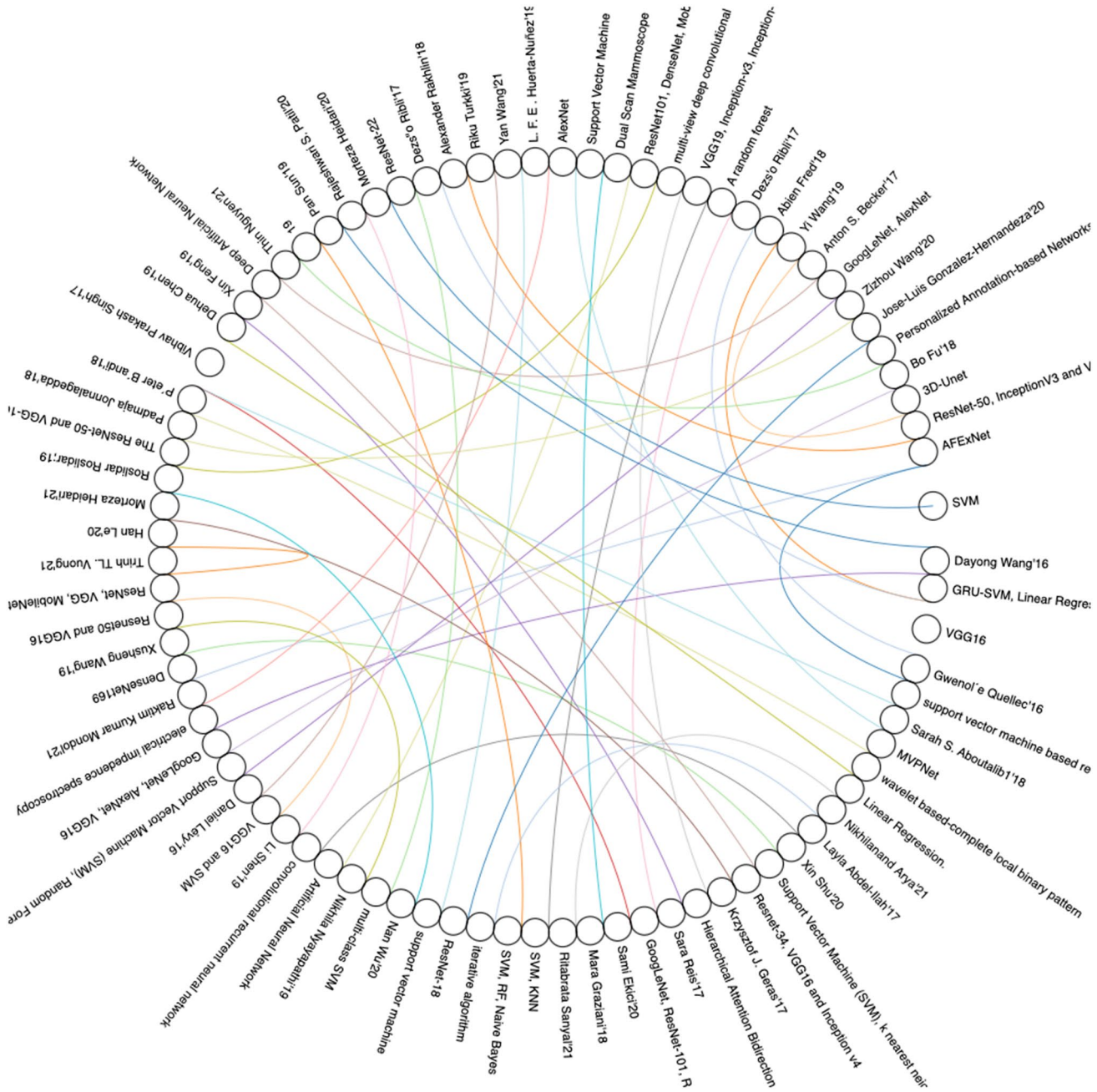


Fig. 12 Chord chart showing the connection of author with deep learning and machine learning techniques from last 5 years

The breast cancer detection techniques are broadly categorized into four classes such as image enhancement, lesion segmentation, feature extraction, and classification techniques (see Fig. 11). Deep learning techniques are used in fourth category, i.e., classification. Figure 12 shows the connection of authors with the deep learning techniques that are used for breast cancer detection from last 5 years.

Wang et al. [53] used ResNet-50 to detect breast lesion regions. They used self-created dataset with the help of West China Hospital. Wang et al. [54] created a dataset using the Invenia ABUS system in Sun Yat-Sen University Cancer Center. They used Unet and DDS pooling to enhance the detection sensitivity for breast cancer. The sensitivity obtained from their model was 0.95. Shu et al. [55] used INbreast dataset and CBIS dataset for breast cancer detection. Pre-trained CNNs and deep machine learning models were used for classification of breast lesion region. Heidari et al. [4] used SVM classifier for breast image. Fu et al. [100] proposed a model to predict Invasive Disease-Free Survival (iDFS) for the early-stage breast cancer patients. They used

XGBoosting and attained the AUC of 0.845. The dataset was self-created with the help of CRCB in West China Hospital of Sichuan University. Nyayapathi et al. [76] shown the fusion of ultrasound images with photoacoustic images for the early detection of breast cancer. Experimental results revealed the improvement in probability, speed, and also the patient's comfort zone. Ekici et al. [51] used CNN classifier on DMR dataset and obtained the accuracy of 98.95%.

Yin et al. [34] proposed a new scheme for the early detection using UWCEM database. They created 3-D anatomically accurate FDTD-based breast models, which resulted in improved identification capability, robust artifact resistance, and high detectability of tumors. Srivastava et al. [81] used different SVM classifiers (i.e., MLP, Quadratic, linear, and RBF) to achieve good accuracy. SVM-RBF performed better than the others. SVM-RBF attained the accuracy of 87.5%. Dheeba et al. [101] worked on the detection of tumors from breast tissue structure using a mammogram. MIAS dataset and ANN classifier were used. The recognition score obtained from this method was 97.8%. Wang et al. [53] used SoftMax classifier on DDSM dataset to detect the size of breast lesions. The value of AUC obtained from this method was 0.865. However, there is a need to develop more advance automatic breast cancer detection techniques to improve the diagnosis of breast cancer. Table 11 shows the advancement in breast cancer detection techniques.

6.3 Comparative analysis

The effectiveness of deep learning-based breast cancer detection techniques is assessed using the performance measures mentioned in Sect. 4. Table 12 shows the comparative analysis of deep learning and machine learning techniques in terms of performance measures. In this table, denotes that the approach has used the associated performance meaures, while indicates otherwise. The quantative analysis of different deep learning and machine learning techniques is illustrated in Table 13.

7 Discussion

In order to assist in cancer treatment, diagnostic imaging modalities are important for tumor classification. Over the last few years, imaging is considered an important tool for the diagnosis of tumors. Various screening techniques are used to detect and characterize the tumors. Screening techniques do not prevent cancer, however, they make early detection possible to make the patient alert for their treatment. Every screening technique has its own advantages and disadvantages. The most common screening method for breast cancer detection is an X-ray mammogram. Breast Ultrasound creates an image of tissues inside the breast using sound waves. The advantages involve non-invasive, quick visualization of breast tissue, the area closest to the chest wall, which is difficult to study with a mammogram, can be seen with ultrasound. Breast ultrasound has a low specificity and is more expensive. MRIs use a large magnet and radio waves to produce excellent tissue differentiation and sensitivity for breast cancer detection. CT scan images are made up of X-rays taken from various angles. Patients look for non-invasive and non-contact screening techniques as all the above-mentioned techniques involve contact. Thermography is one such technique. It uses infrared sensors to detect heat and increased vascularity as the result of biochemical reactions. Table 14 depicts the comparison of different breast tumor screening techniques in terms of pros and cons.

The performance of system can be judged by its accuracy, sensitivity, and specificity. The value of these measures should be high to achieve good results. When the likelihood of malignancy is established using a system, lively observation or biopsy can be recommended to evade inadequate and further invasive treatment. The main focus of this study is multi-scale module to improve the performance.

7.1 Relevance of Clinical Validation

Cancer severity can be measured through cancer research by describing a normal or abnormal state of cancer in the individual [113–115]. For this, the performance measures of cancer detection techniques are analyzed. In terms of clinical relevance, a cancer detection study may assess the risk of developing cancer in a specific tissue, or it may assess the risk of cancer progression or response to therapy. The conceptual framework of cancer research is mentioned in Fig. 13. The framework involves multiple processes, linking initial discovery in basic studies, validation, and clinical implementation.

A technique is discovered at the start of any cancer research technique development, and it is typically validated within the same initial report. When independent patient sets are not available, cross-validation-based methods are frequently used to replace validation based on a predefined prediction rule in an independent patient series. Prior to the analysis, the research question and plan, as well as the fundamental use of the technique, research design is clearly defined. Analytical validation is carried out after the new cancer research development phase [116]. This demonstrates how accurately and consistently the test measures the patient's analyte(s). Next is clinical validation, it tests to confirm its ability to predict or diagnose the clinical phenotype or outcome of interest, as demonstrated during the discovery and initial validation phases. A cancer detection

Table 11 Advancement in breast cancer detection techniques

References	Proposed approach	Targeted problem	Models	Classifier	Measures	Challenges	Advantages
[53]	Create dataset + Multiscale module (screening of unique features) + Multi-instance module	Detection of the breast lesion regions	CNN (ResNet-50)	–	AUC: ResNet-50 MSMI = 0.901	In future author could assist other advanced applications, such as size measurement, lesion characterization	Their method learned the unique features of lesions via the multi-scale module
[54]	Pre-processing + pre-trained UNet + DDS feature extraction + TM	Enhance the detection sensitivity for breast cancer	3D CNN (U-net + DDS Pool)	–	Sensitivity = 95% with 0.84 FP per/volume	Author could help with more advanced applications in the future, such as size measurement and lesion characterization	High sensitivity and low false positives
[55]	Edge detection and segmentation + feature extraction + RGP/GGP	Classification based on breast lesion region	CNN (DenseNet169 + max pooling)	Pretrained CNNs and deep MIL models	Acc: INbreast dataset = 0.934 ± 0.0003(RGP) 0.922 ± 0.0002(GGP) CBIS dataset = 0.838 ± 0.0001(RGP) 0.767 ± 0.0002(GGP)	The results of visualization show that the proposed model can roughly locate suspicious regions	Ability of learning lesion location information
[71]	Data Preprocessing + Image feature extraction + Classification	Classification based on image feature extraction	CNN (SVM)	SVM (FFDM image, DCT maps, FFT maps, fusion of features)	FFT: 66–77% DCT: 64–83% SSIM: 63–71% Fusion: 67–89%	Its clinical utility or impact on radiologists' performance in diagnosis of breast cancer using mammograms has not been tested	significantly higher performance with AUC
[100]	Statistical Feature Selection + Ensemble Feature Selection + XGBoost algorithm + PREDICT and Adjuvant Online	Predict the relapse or metastasis breast cancer	ANN (XGBoost)	PREDICT and Adjuvant Online	AUC: 0.8451	In future works, additional clinical data can be collected for improving accuracy of the 5-year iDFS prediction algorithm	XGBoost algorithm performance improved by 3%
[76]	System Design + Breast Coupling + Imaging Procedure + Reconstruction and Alignment	Early detection of tumor	–	Fusion of ultrasound images with photoacoustic images	benefits of portability, speedy scanning, and patient comfort	Patients with different tumor characteristics and breast sizes should be imaged to identify the photoacoustic features of different tumor grades and types	Their system possesses the benefits of portability, speedy scanning, and patient comfort

Table 11 (continued)

References	Proposed approach	Targeted problem	Models	Classifier	Measures	Challenges	Advantages
[51]	Data acquisition + Preprocessing + Segmentation + Feature Extraction + Classification	Classification based on image feature extraction	CNN	CNN (Bayes algorithm)	Accuracy: 98.95%	The author can further classify the type and size of tumor	Obtained good accuracy rate was obtained for the thermal images
[87]	Digital breast model + Parameter estimation for thermal images	Detect and localize tumor		iCAD	–	Requires more computation time	The technique has potential to be an accurate adjunct to mammography
[75]	Data acquisition + designed deep model + cross domain SVM	learning deep features from large-scale X-ray images	CNN	CNN (multi-class SVM)		Author could help with other advanced applications in the future, such as size measurement and lesion characterization	Good effectiveness and efficiency of proposed recognition system
[93]	Create dataset + Model-Based Segmentation + Compute optical flow	Accurately measuring 3-D surface motion	–	DIET machine	The system can detect a 10 mm tumor in a silicone phantom breast	A limitation of these dense optical flow techniques is their sensitivity to lighting variation	Their system can reconstruct the breast surface with average errors of less than 1 mm
[48]	Heterogeneity Wavelet Kinetic Features + Classification	Quantify intra-tumor heterogeneity in breast cancer	–	Dynamic contrast-enhanced magnetic resonance imaging	AUCs: 0.94	HetWave could assist other advanced applications such as feature extraction approach for characterizing tumor heterogeneity, providing valuable prognostic information	Superior ROC AUC
[34]	3-D anatomically accurate FDTD-based breast models + Preprocessing for Artifact Removal + RAR algorithm	Early stage detection	Robust and Artifact Resistant	DAS, DMAS, MWDAS, and FDAS	Improved identification capability, robust artifact resistance, and high detectability	The investigation of RAR's performance for further enhancement of tumor detection in severely dense breasts is missing	Their results show the high potential of RAR for the early stage cancer detection in low to medium density breasts
[102]	Breast-Body Segmentation	Automatically compute breast density in breast MRI	–			Complex analysis	Their proposed method is good at investigating the correlation between breast density measurements obtained from MRI and mammograms

Table 11 (continued)

References	Proposed approach	Targeted problem	Models	Classifier	Measures	Challenges	Advantages
[2]	Input raw pixels of training patches + Feature representation (SSAE) + Softmax classifier	Appearance of breast cancer nuclei in histopathological images	SSAE	Softmax Classifier	SSAE + SMC: Precision = 88.84% F-measure = 84.49% AveP = 78.83%	Stacked Sparse Autoencoder could be integrated with feature extraction methods to characterize cancer better	Their framework can provide accurate seed points or vertices for developing cell-by-cell graph features that can enable characterization of cellular topology features on tumor histology
[81]	Data Acquisition + Segmentation + Feature extraction and selection + Classifier	Feature extraction	ANN (GA-MI)	SVM (MLP, Quadratic, linear, RBF)	SVM-MLP: accuracy 84.3750% SVM-linear: accuracy 81.25% SVM-quadratic: accuracy 84.3750% SVM-RBF: accuracy 87.5%	Some SVM classifier's are not capable of classifying the negative sample	GA-MI based feature selection the SVM classifiers for MLP, linear, and quadratic classifier are performing better
[101]	Data acquisition + ROI Selection + Feature extraction + classification	Detection of tumor from breast tissue structure using mammogram	ANN	ANN (Modified Genetic Algorithm)	Recognition score of 97.8%	The results are not comparatively good on other mammogram datasets	Classifier is good at recognition
[103]	Data acquisition + slide level preprocessing + ResNet Classifier	cancer metastases in lymph nodes	CNN	ResNet	The best results were obtained with pre-trained architectures such as ResNet	At the slide level, the best ranked team misclassified 67 of the 500 slides in the test set	The submitted algorithms were not only able to detect the presence of metastases but also measure their extent to derive the metastasis category, including ITC, and to determine the pN-stage that is used in clinical practice

Table 12 Comparative analysis of braest cancer detection techniques in terms of performance measures

References	Technique	Prediction measure					Classification measure				
		μ	σ	*MSE*	*RMSE*	*PSNR*	*PPV*	S_n	A_c	S_f	*AUC-ROC*
[104]	Resnet-34, VGG16	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
[105]	Resnet50, VGG16	✗	✗	✗	✗	✗	✗	✓	✗	✓	✓
[50]	VGG16	✗	✗	✗	✗	✗	✗	✓	✗	✓	✓
[106]	Linear regression	✗	✗	✗	✗	✗	✗	✓	✓	✗	✗
[54]	3D-Unet	✗	✗	✗	✗	✗	✗	✓	✗	✗	✗
[55]	DenseNet169	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
[107]	MVPNet	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
[108]	EfficientNet	✗	✗	✗	✗	✗	✗	✓	✓	✓	✓
[109]	AFExNet	✓	✓	✗	✓	✗	✓	✓	✓	✓	✓
[110]	GoogLeNet, AlexNet	✓	✗	✗	✗	✗	✗	✓	✓	✗	✓
[111]	Random forest	✗	✗	✗	✗	✗	✗	✗	✓	✗	✓
[112]	DTree, RF, XGBoost	✗	✗	✗	✗	✗	✗	✓	✓	✓	✗

Table 13 Quantitative assessment of breast cancer prediction technique on classification measures

References	Technique	Classification measures				
		PPV	S_n	A_c	S_f	AUC-ROC
[104]	Resnet-34, VGG16	–	–	89.0%	–	0.950
[105]	Resnet50, VGG16	–	86.7%	–	96.1%	0.98
[50]	VGG16	–	0.9	–	–	085
[106]	Linear Regression	–	–	92.43 ± 0.657	–	–
[54]	3D–Unet	0.84	95%	–	–	–
[55]	DenseNet169	–		0.923 ± 0.0003 0.762 ± 0.0002		0.762 ± 0.0002 0.838 ± 0.0001
[107]	MVPNet	–	$94.2 \pm 2.2\%$	92.2%	$92.3 \pm 2.4\%$	0.91 = /–0.05
[108]	EfficientNet	0.819	0.74	90.2%	95%	0.93
[109]	AFExNet	98.57	98.58	98.57	98.57	–
[110]	GoogLeNet, AlexNet	0.7051	–	–	–	0.925
[111]	Random forest	–	–	84%	–	0.84–0.86
[112]	DTree, RF, XGBoost	–	0.8429	86.96%	0.8964	–

technique that has been analytically and clinically validated is now ready for use in clinical care.

8 Future Research Directions

Many females overlook the need of breast cancer diagnosis because they find it discomforting. Most of the breast cancer screening techniques are invasive. An alternative technique is required for breast cancer screening in which the skin is not pierced. The possible future research directions are as follows:

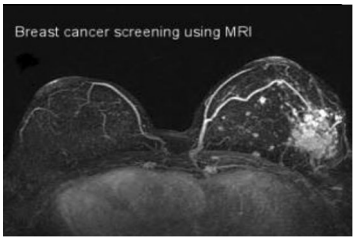
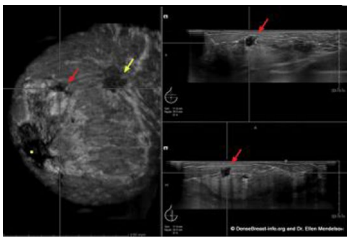
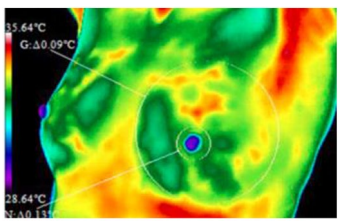
8.1 Ensure Safe Engagement

Safety is a major issue when radioactive rays or ionizing radiation are directly exposed to one's body. Many screening techniques use harmful waves that may cause the allergic reaction or contraindicate in some patients [117–119]. A system should be designed in such a way that it uses non-ionizing radiations and also provides accurate results.

8.2 Multimodal Approaches

Multimodality-based approach should be considered for the detection of breast cancer at an early stage. Screening tools for breast cancer need to expand their expertise by providing the multimodal method to enhance the precision by improving the outcome of screening techniques.

Table 14 Comparison of different breast tumor screening techniques for diagnosis of tumor

Screening Technique	Spatial Resolution	Advantages	Disadvantages	Image
Mammogram	Lower Spatial resolution	Time efficient, requires lower average radiation dosage	Lower spatial resolution, uses X-rays, costly	
CT Scan	High spatial resolution	Precise, High spatial resolution,	Uses X-rays, can cause allergic reaction	
MRI	High spatial resolution	No-ionizing radiation, high sensitivity, good tissue differentiation	Costly, invasive procedure, low specificity	
Ultrasound	Moderate spatial resolution	Less expensive, uses soundwaves	Low specificity, not able to detect all types of tumors	
Thermography	Low spatial resolution	Non-invasive, does not involve exposure to radiation	It can only alert a person to changes that may need further investigation	

8.3 Model Generalization

A variety of deep learning models are used in breast cancer research. Deep learning models provide different results for different application site [120, 121]. Hence, there is a necessity to develop the generalized model for breast cancer detection.

8.4 Clinical Implementation

Deep learning-based breast cancer detection models proved their significance in the medical research. However, the practical implementation of these models in clinics is still not done [122]. The implementation of these model in clincis will be beneficial for doctors.

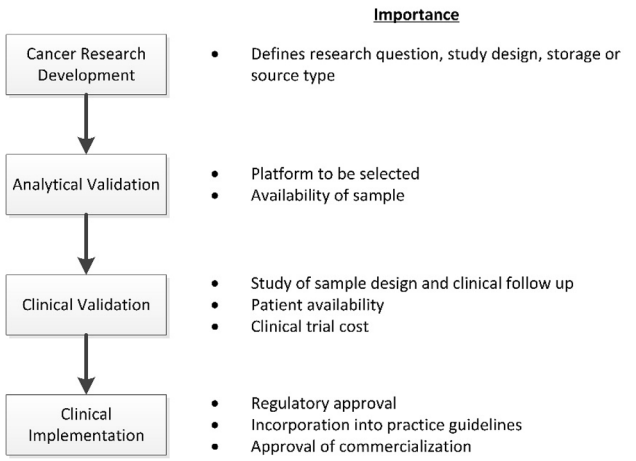


Fig. 13 Conceptual framework of cancer research

9 Conclusions

There are several methods for diagnosis and planning of initial breast cancer screening such as mammography, CT, breast ultrasound, MRI, and thermography. The outcome of these screening procedures help the doctors to aid in the selection of treatment or track the status of recovery. The aim of this paper is to provide the useful information to recognize and classify the breast tumor to make the early detection so that life can be saved. An attempt has been made to investigate the breast cancer detection using different intelligent systems along with their datasets. In addition to recent developments in various imaging techniques, the problems associated with the existing techniques have also been discussed. The possible future research directions for the ideal imaging modality are also suggested. In the wake of the limitations of existing techniques, it is imperative to improve the existing techniques for breast cancer detection as the core of personalized medicine and healthcare remains to determine the most suitable screening modality for the proper diagnosis of breast cancer.

Declarations

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

References

1. Bahramiabarghouei H, Porter E, Santorelli A, Gosselin B, Popovíc M, Rusch LA (2015) Flexible 16 antenna array for microwave breast cancer detection. IEEE Trans Biomed Eng 62(10):2516–2525. https://doi.org/10.1109/TBME.2015.2434956
2. Xu J et al (2016) Stacked sparse autoencoder (SSAE) for nuclei detection on breast cancer histopathology images. IEEE Trans Med Imaging 35(1):119–130. https://doi.org/10.1109/TMI.2015.2458702
3. National Cancer Registry Programme (2008) Report of population based cancer registries 2012–2018 National Cancer Registry Programme. National Centre for Disease Informatics and Research. and Indian Council of Medical Research, Bangalore. https://www.ncdirindia.org/Reports.aspx. Accessed 3 Dec 2020
4. Eismann J et al (2019) Interdisciplinary management of transgender individuals at risk for breast cancer: case reports and review of the literature. Clin Breast Cancer 19(1):e12–e19. https://doi.org/10.1016/j.clbc.2018.11.007
5. Stone JP, Hartley RL, Temple-Oberle C (2018) Breast cancer in transgender patients: a systematic review. Part 2: female to male. Eur J Surg Oncol. https://doi.org/10.1016/j.ejso.2018.06.021
6. Gooren LJ, van Trotsenburg MAA, Giltay EJ, van Diest PJ (2013) Breast cancer development in transsexual subjects receiving cross-sex hormone treatment. J Sex Med 10(12):3129–3134. https://doi.org/10.1111/jsm.12319
7. De Blok CJM et al (2019) Breast cancer risk in transgender people receiving hormone treatment: Nationwide cohort study in the Netherlands. BMJ. https://doi.org/10.1136/bmj.l1652
8. Nikolic DV et al (2012) Importance of revealing a rare case of breast cancer in a female to male transsexual after bilateral mastectomy. World J Surg Oncol 10:2–5. https://doi.org/10.1186/1477-7819-10-280
9. Chen D, Huang M, Li W (2019) Knowledge-powered deep breast tumor classification with multiple medical reports. IEEE/ACM Trans Comput Biol Bioinf. https://doi.org/10.1109/tcbb.2019.2955484
10. Duffy SW et al (2002) The impact of organized mammography service screening on breast carcinoma mortality in seven Swedish counties: a collaborative evaluation. Cancer 95(3):458–469. https://doi.org/10.1002/cncr.10765
11. Watine J (2002) "Prognostic factors for patients with small lung carcinoma. Cancer 94(2):576–578. https://doi.org/10.1002/cncr.10243
12. Duffy S, Tabár L, Smith RA (2002) The mammographic screening trials: commentary on the recent work by Olsen and Gøtzsche. J Surg Oncol 81(4):68–71. https://doi.org/10.1002/jso.10193
13. Ferlay J et al (2015) Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. Int J cancer 136(5):E359–E386
14. US Preventive Services Task Force (2002) Clinical Guidelines Preventive Services Task Force. Ann Intern Med 137(11):917–933
15. Boyd NF et al (2006) Body size, mammographic density, and breast cancer risk. Cancer Epidemiol Biomarkers Prev 15(11):2086–2092. https://doi.org/10.1158/1055-9965.EPI-06-0345
16. Carney PA, Miglioretti DL, Yankaskas BC, Kerlikowske K, Rosenberg R, Rutter CM (2003) Erratum: Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography (Annals of Internal Medicine (2003) 138 (168–175)). Ann Intern Med 138(9):771. https://doi.org/10.7326/0003-4819-138-9-200305060-00025
17. Elter M, Horsch A (2009) CADx of mammographic masses and clustered microcalcifications: a review. Med Phys 36(6):2052–2068. https://doi.org/10.1118/1.3121511
18. Marmot MG, Altman DG, Cameron DA, Dewar JA, Thompson SG, Wilcox M (2013) The benefits and harms of breast cancer

screening: an independent review. Br J Cancer 108(11):2205–2240. https://doi.org/10.1038/bjc.2013.177
19. Gu X et al (2020) Age-associated genes in human mammary gland drive human breast cancer progression. Breast Cancer Res 22(1):1–15. https://doi.org/10.1186/s13058-020-01299-2
20. Ferguson NL et al (2013) Prognostic value of breast cancer subtypes, Ki-67 proliferation index, age, and pathologic tumor characteristics on breast cancer survival in caucasian women. Breast J 19(1):22–30. https://doi.org/10.1111/tbj.12059
21. Rana P, Ratcliffe J, Sussman J, Forbes M, Levine M, Hodgson N (2017) Young women with breast cancer: needs and experiences. Cogent Med 4(1):1–11. https://doi.org/10.1080/2331205x.2017.1278836
22. Anderson BO et al (2008) Guideline implementation for breast healthcare in low-income and middle-income countries: Overview of the breast health global initiative Global Summit 2007. Cancer 113(8 Suppl.):2221–2243. https://doi.org/10.1002/cncr.23844
23. Hopp T, Duric N, Ruiter NV (2015) Image fusion of Ultrasound Computer Tomography volumes with X-ray mammograms using a biomechanical model based 2D/3D registration. Comput Med Imaging Graph 40:170–181. https://doi.org/10.1016/j.compmedimag.2014.10.005
24. Pisano ED et al (2008) Diagnostic accuracy of digital versus film mammography: Exploratory analysis of selected population subgroups in DMIST. Radiology 246(2):376–383. https://doi.org/10.1148/radiol.2461070200
25. Yassin NIR, Omran S, El Houby EMF, Allam H (2018) Machine learning techniques for breast cancer computer aided diagnosis using different image modalities: a systematic review. Comput Methods Programs Biomed 156:25–45. https://doi.org/10.1016/j.cmpb.2017.12.012
26. Gupta NP, Malik PK, Ram BS (2020) A review on methods and systems for early breast cancer detection. In: Proceedings of international conference on computation, automation and knowledge management (ICCAKM 2020), 42–46. https://doi.org/10.1109/ICCAKM46823.2020.9051554
27. Lu Y, Li JY, Su YT, Liu AA (2018) A review of breast cancer detection in medical images. In: VCIP 2018—IEEE international conference on visual communications and image processing conference, pp 11–14. https://doi.org/10.1109/VCIP.2018.8698732
28. Huppe AI, Mehta AK, Brem RF (2018) Molecular breast imaging: a comprehensive review. Semin Ultrasound CT MRI 39(1):60–69. https://doi.org/10.1053/j.sult.2017.10.001
29. Oyelade ON, Ezugwu AES (2020) A state-of-the-art survey on deep learning methods for detection of architectural distortion from digital mammography. IEEE Access 8:148644–148676. https://doi.org/10.1109/ACCESS.2020.3016223
30. Al Husaini MAS, Habaebi MH, Hameed SA, Islam MR, Gunawan TS (2020) A systematic review of breast cancer detection using thermography and neural networks. IEEE Access 8:208922–208937. https://doi.org/10.1109/ACCESS.2020.3038817
31. Hartley RL, Stone JP, Temple-Oberle C (2018) Breast cancer in transgender patients: a systematic review. Part 1: male to female. Eur J Surg Oncol. https://doi.org/10.1016/j.ejso.2018.06.035
32. Page MJ et al (2021) The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. BMJ 372(71):1–9
33. Agarwal T, Kumar V (2021) A systematic review on bat algorithm: theortical foundations, variants, and applications. Archives of Computational Methods in Engineering. https://doi.org/10.1007/s11831-021-09673-9
34. Yin T, Ali FH, Reyes-Aldasoro CC (2015) A robust and artifact resistant algorithm of ultrawideband imaging system for breast cancer detection. IEEE Trans Biomed Eng 62(6):1514–1525. https://doi.org/10.1109/TBME.2015.2393256
35. Li Q et al (2015) Direct extraction of tumor response based on ensemble empirical mode decomposition for image reconstruction of early breast cancer detection by UWB. IEEE Trans Biomed Circuits Syst 9(5):710–724. https://doi.org/10.1109/TBCAS.2015.2481940
36. Alexandrou G et al (2021) Detection of multiple breast cancer ESR1 mutations on an ISFET based Lab-on-chip platform. IEEE Trans Biomed Circuits Syst 15(3):380–389. https://doi.org/10.1109/TBCAS.2021.3094464
37. Suckling J. The mini-MIAS database of mammograms. http://peipa.essex.ac.uk/info/mias.html
38. Kopans D. DDSM: digital database for screening mammography. http://www.eng.usf.edu/cvprg/Mammography/Database.html. Accessed 24 Dec 2020
39. Moreira IC, Amaral I, Domingues I, Cardoso A, Cardoso MJ, Cardoso JS (2012) Inbreast: toward a full-field digital mammographic database. Acad Radiol 19(2):236–248
40. Prapavesis S, Fornage BD, Weismann CF, Palko A, Zoumpoulis P (2003) Breast ultrasound and US-guided interventional techniques: a multimedia teaching file. Thessaloniki, Greece
41. Yap MH et al (2018) Automated breast ultrasound lesions detection using convolutional neural networks. IEEE J Biomed Health Inf 22(4):1218–1226
42. Al-Dhabyani W, Gomaa M, Khaled H, Fahmy A (2020) Dataset of breast ultrasound images. Data Br 28:104863
43. Man Ng HH et al (2019) Hepatitis B virus-associated intrahepatic cholangiocarcinoma has distinct clinical, pathological and radiological characteristics: a systematic review. Surg Gastroenterol Oncol 24(1):5. https://doi.org/10.21614/sgo-24-1-5
44. Commean P. RIDER breast MRI. https://wiki.cancerimagingarchive.net/display/Public/RIDER+Breast+MRI
45. Kirby J. QIN breast DCE-MRI. https://wiki.cancerimagingarchive.net/display/Public/QIN+Breast+DCE-MRI
46. Silva LF et al (2014) A new database for breast research with infrared image. J Med Imaging Heal Inf 4(1):92–100. https://doi.org/10.1166/jmihi.2014.1226
47. Bhowmik MK, Gogoi UR, Majumdar G, Bhattacharjee D, Datta D, Ghosh AK (2017) Designing of ground-truth-annotated DBT-TU-JU breast thermogram database toward early abnormality prediction. IEEE J Biomed Heal Inf 22(4):1238–1249
48. Mahrooghy M et al (2015) Pharmacokinetic tumor heterogeneity as a prognostic biomarker for classifying breast cancer recurrence risk. IEEE Trans Biomed Eng 62(6):1585–1594. https://doi.org/10.1109/TBME.2015.2395812
49. Singh VP, Srivastava S, Srivastava R (2017) Effective mammogram classification based on center symmetric-LBP features in wavelet domain using random forests. Technol Health Care 25(4):709–727. https://doi.org/10.3233/THC-170851
50. Ribli D, Horváth A, Unger Z, Pollner P, Csabai I (2018) Detecting and classifying lesions in mammograms with Deep Learning. Sci Rep 8(1):16–20. https://doi.org/10.1038/s41598-018-22437-z
51. Ekici S, Jawzal H (2020) Breast cancer diagnosis using thermography and convolutional neural networks. Med Hypotheses 137:109542. https://doi.org/10.1016/j.mehy.2019.109542
52. Roslidar R, Saddami K, Arnia F, Syukri M, Munadi K (2019) A study of fine-tuning CNN models based on thermal imaging for breast cancer classification. In: Proceedings of Cybernetics 2019—2019 International conference on cybernetics and computational intelligence: towards a smart and human-centered cyber world, pp 77–81. https://doi.org/10.1109/CYBERNETICSCOM.2019.8875661
53. Wang Z, Zhang L, Shu X, Lv Q, Yi Z (2020) An end-to-end mammogram diagnosis: a new multi-instance and multi-scale method based on single-image feature. IEEE Trans Cogn Dev Syst. https://doi.org/10.1109/TCDS.2019.2963682

54. Wang Y et al (2020) Deeply-supervised networks with threshold loss for cancer detection in automated breast ultrasound. IEEE Trans Med Imaging 39(4):866–876. https://doi.org/10.1109/TMI.2019.2936500
55. Shu X, Zhang L, Wang Z, Lv Q, Yi Z (2020) Deep neural networks with region-based pooling structures for mammographic image classification. IEEE Trans Med Imaging 39(6):2246–2255. https://doi.org/10.1109/TMI.2020.2968397
56. Dogra N, Kumar V (2021) A comphrehensive review on deep synergistic drug prediction techniques for cancer. Arch Comput Methods Eng 29:1443–1461
57. Helwan A, Abiyev R (2016) Shape and texture features for the identification of breast cancer. Lecture notes in computational science and engineering, vol 2226, pp 542–547
58. Pandit VR, Bhiwani RJ (2015) Image fusion in remote sensing applications: a review. Int J Comput Appl 120(10):22–32. https://doi.org/10.5120/21263-3846
59. El Hami A, Pougnet P (2019) Embedded mechatronic system 2: analyses of failures, modeling, simulation and optimization. Elsevier, Amsterdam, p 300
60. Yang Y, Wan W, Huang S, Yuan F, Yang S, Que Y (2016) Remote sensing image fusion based on adaptive IHS and multiscale guided filter. IEEE Access 4:4573–4582. https://doi.org/10.1109/ACCESS.2016.2599403
61. Naidu VPS, Raol JR (2008) Pixel-level image fusion using wavelets and principal component analysis. Def Sci J 58(3):338–352. https://doi.org/10.14429/dsj.58.1653
62. Murtaza G et al (2020) Deep learning-based breast cancer classification through medical imaging modalities: state of the art and research challenges. Artif Intell Rev 53(3):1655–1720. https://doi.org/10.1007/s10462-019-09716-5
63. Agarwal S (2014) Data mining: data mining concepts and techniques. In: 2013 International conference on machine intelligence and research advancement (ICMIRA)
64. Edman MC, Marchelletta RR, Hamm-Alvarez SF (2010). Lacrimal gland overview. In: Encyclopedia of the eye, pp 522–527. https://doi.org/10.1016/B978-0-12-374203-2.00049-X.
65. Barnes NLP, Ooi JL, Yarnold JR, Bundred NJ (2012) Ductal carcinoma in situ of the breast. BMJ 344(7846):1430–1441. https://doi.org/10.1136/bmj.e797
66. Schnitt SJ (2010) Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. Mod Pathol 23:60–64. https://doi.org/10.1038/modpathol.2010.33
67. Gupta A, Shridhar K, Dhillon PK (2015) A review of breast cancer awareness among women in India: cancer literate or awareness deficit? Eur J Cancer 51(14):2058–2066. https://doi.org/10.1016/j.ejca.2015.07.008
68. Brooks AD et al (2009) Modern breast cancer detection: a technological review. Int J Biomed Imaging. https://doi.org/10.1155/2009/902326
69. Nam KJ et al (2015) Comparison of full-field digital mammography and digital breast tomosynthesis in ultrasonography-detected breast cancers. Breast 24(5):649–655. https://doi.org/10.1016/j.breast.2015.07.039
70. Abdelhafiz D, Yang C, Ammar R, Nabavi S (2019) Deep convolutional neural networks for mammography: advances, challenges and applications. BMC Bioinf. https://doi.org/10.1186/s12859-019-2823-4
71. Heidari M, Mirniaharikandehei S, Liu W, Hollingsworth AB, Liu H, Zheng B (2020) Development and assessment of a new global mammographic image feature analysis scheme to predict likelihood of malignant cases. IEEE Trans Med Imaging 39(4):1235–1244. https://doi.org/10.1109/TMI.2019.2946490
72. Shen S et al (2015) A multi-centre randomised trial comparing ultrasound vs mammography for screening breast cancer in high-risk Chinese women. Br J Cancer 112(6):998–1004. https://doi.org/10.1038/bjc.2015.33
73. Brem RF, Lenihan MJ, Lieberman J, Torrente J (2015) Screening breast ultrasound: past, present, and future. Am J Roentgenol 204(2):234–240. https://doi.org/10.2214/AJR.13.12072
74. Kolb TM, Lichy J, Newhouse JH (2002) Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology 225(1):165–175. https://doi.org/10.1148/radiol.2251011667
75. Wang X, Chen X, Cao C (2019) Hierarchically engineering quality-related perceptual features for understanding breast cancer. J Vis Commun Image Represent 64:102644. https://doi.org/10.1016/j.jvcir.2019.102644
76. Nyayapathi N et al (2020) Dual scan mammoscope (DSM)—a new portable photoacoustic breast imaging system with scanning in craniocaudal plane. IEEE Trans Biomed Eng 67(5):1321–1327. https://doi.org/10.1109/TBME.2019.2936088
77. Mann RM, Cho N, Moy L (2019) Reviews and commentary—state of the art. Radiology 292(3):520–536. https://doi.org/10.1148/radiol.2019182947
78. Sun P, Wang D, Mok VC, Shi L (2019) Comparison of feature selection methods and machine learning classifiers for radiomics analysis in glioma grading. IEEE Access 7:102010–102020. https://doi.org/10.1109/access.2019.2928975
79. Wu N et al (2020) Deep neural networks improve radiologists' performance in breast cancer screening. IEEE Trans Med Imaging 39(4):1184–1194. https://doi.org/10.1109/TMI.2019.2945514
80. Kale MC, Clymer BD, Koch RM, Heverhagen JT, Sammet S, Stevens R, Knopp MV (2008) Multispectral co-occurrence with three random variables in dynamic contrast enhanced magnetic resonance imaging of breast cancer. IEEE Trans Med Imaging 27(10):1425–1431
81. Srivastava S, Sharma N, Singh SK, Srivastava R (2014) Quantitative analysis of a general framework of a CAD tool for breast cancer detection from mammograms. J Med Imaging Heal Inf 4(5):654–674. https://doi.org/10.1166/jmihi.2014.1304
82. Haeri Z, Shokoufi M, Jenab M, Janzen R, Golnaraghi F (2016) Electrical impedance spectroscopy for breast cancer diagnosis: clinical study. Integr. Cancer Sci Ther 3(6):1–6. https://doi.org/10.1038/nprot.2014.110
83. Huerta-Nuñez LFE et al (2019) A biosensor capable of identifying low quantities of breast cancer cells by electrical impedance spectroscopy. Sci Rep 9(1):1–12. https://doi.org/10.1038/s41598-019-42776-9
84. Lederman D, Zheng B, Wang X, Wang XH, Gur D (2011) Improving breast cancer risk stratification using resonance-frequency electrical impedance spectroscopy through fusion of multiple classifiers. Ann Biomed Eng 39(3):931–945. https://doi.org/10.1007/s10439-010-0210-4
85. Ward LC, Dylke E, Czerniec S, Isenring E, Kilbreath SL (2011) Confirmation of the reference impedance ratios used for assessment of breast cancer-related lymphedema by bioelectrical impedance spectroscopy. Lymphat Res Biol 9(1):47–51. https://doi.org/10.1089/lrb.2010.0014
86. Etehadtavakol M, Chandran V, Ng EYK, Kafieh R (2013) Breast cancer detection from thermal images using bispectral invariant features. Int J Therm Sci 69:21–36. https://doi.org/10.1016/j.ijthermalsci.2013.03.001
87. Gonzalez-Hernandez JL, Recinella AN, Kandlikar SG, Dabydeen D, Medeiros L, Phatak P (2020) An inverse heat transfer approach for patient-specific breast cancer detection and tumor localization using surface thermal images in the prone position. Infrared Phys Technol 105:103202. https://doi.org/10.1016/j.infrared.2020.103202

88. Mambou SJ, Maresova P, Krejcar O, Selamat A, Kuca K (2018) Breast cancer detection using infrared thermal imaging and a deep learning model. Sensors (Switzerland). https://doi.org/10.3390/s18092799
89. Casalegno F et al (2019) Caries detection with near-infrared transillumination using deep learning. J Dent Res 98(11):1227–1233. https://doi.org/10.1177/0022034519871884
90. Klemm M, Leendertz JA, Gibbins D, Craddock IJ, Preece A, Benjamin R (2009) Microwave radar-based breast cancer detection: Imaging in inhomogeneous breast phantoms. IEEE Antennas Wirel Propag Lett 8:1349–1352. https://doi.org/10.1109/LAWP.2009.2036748
91. Grzegorczyk TM, Meaney PM, Kaufman PA, Diflorio-Alexander RM, Paulsen KD (2012) Fast 3-D tomographic microwave imaging for breast cancer detection. IEEE Trans Med Imaging 31(8):1584–1592. https://doi.org/10.1109/TMI.2012.2197218
92. Tunçay AH, Akduman I (2015) Realistic microwave breast models through T1-weighted 3-D MRI data. IEEE Trans Biomed Eng 62(2):688–698. https://doi.org/10.1109/TBME.2014.2364015
93. Botterill T, Lotz T, Kashif A, Chase JG (2014) Reconstructing 3-D skin surface motion for the DIET breast cancer screening system. IEEE Trans Med Imaging 33(5):1109–1118. https://doi.org/10.1109/TMI.2014.2304959
94. Kao TJ et al (2008) Regional admittivity spectra with tomosynthesis images for breast cancer detection: preliminary patient study. IEEE Trans Med Imaging 27(12):1762–1768. https://doi.org/10.1109/TMI.2008.926049
95. Baran P et al (2018) High-resolution X-ray phase-contrast 3-d imaging of breast tissue specimens as a possible adjunct to histopathology. IEEE Trans Med Imaging 37(12):2642–2650. https://doi.org/10.1109/TMI.2018.2845905
96. McKnight AL, Kugel JL, Rossman PJ, Manduca A, Hartmann LC, Ehman RL (2002) MR elastography of breast cancer: preliminary results. Am J Roentgenol 178(6):1411–1417. https://doi.org/10.2214/ajr.178.6.1781411
97. Goddi A, Bonardi M, Alessi S (2012) Breast elastography: a literature review. J Ultrasound 15(3):192–198
98. Landhuis E (2020) Deep learning takes on tumours. Nature 580(7804):551–553. https://doi.org/10.1038/d41586-020-01128-8
99. Bengio Y (2009) Learning deep architectures for AI. Found Trends Mach Learn 2(1):1–27. https://doi.org/10.1561/2200000006
100. Fu B, Liu P, Lin J, Deng L, Hu K, Zheng H (2019) Predicting invasive disease-free survival for early stage breast cancer patients using follow-up clinical data. IEEE Trans Biomed Eng 66(7):2053–2064. https://doi.org/10.1109/TBME.2018.2882867
101. Dheeba J, Selvi ST (2011) A CAD system for breast cancer diagnosis using modified genetic algorithm optimized artificial neural network. Lecture notes in computer science (including subseries: Lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 7076, Part 1, pp 349–357. https://doi.org/10.1007/978-3-642-27172-4_43
102. Gubern-Mérida A, Kallenberg M, Mann RM, Martí R, Karssemeijer N (2015) Breast segmentation and density estimation in breast MRI: a fully automatic framework. IEEE J Biomed Health Inf 19(1):349–357. https://doi.org/10.1109/JBHI.2014.2311163
103. Bándi P et al (2019) From detection of individual metastases to classification of lymph node status at the patient level: the CAMELYON17 challenge. IEEE Trans Med Imaging 38(2):550–560. https://doi.org/10.1109/TMI.2018.2867350
104. Le H et al (2020) Utilizing automated breast cancer detection to identify spatial distributions of tumor-infiltrating lymphocytes in invasive breast cancer. Am J Pathol 190(7):1491–1504. https://doi.org/10.1016/j.ajpath.2020.03.012
105. Shen L, Margolies LR, Rothstein JH, Fluder E, McBride R, Sieh W (2019) Deep learning to improve breast cancer detection on screening mammography. Sci Rep. https://doi.org/10.1038/s41598-019-48995-4
106. Graziani M, Andrearczyk V, Müller H (2018) Regression concept vectors for bidirectional explanations in histopathology. Lecture notes in computer science (including subseries: Lecture notes in artificial intelligence and lecture notes in bioinformatics), vol 11038, pp 124–132. https://doi.org/10.1007/978-3-030-02628-8_14
107. Jonnalagedda P, Schmolze D, Bhanu B (2018) MVPNets: multi-viewing path deep learning neural networks for magnification invariant diagnosis in breast cancer. In: Proceedings of 2018 IEEE 18th international conference on bioinformatics and bioengineering (BIBE 2018), pp 189–194. https://doi.org/10.1109/BIBE.2018.00044
108. Arya N, Saha S (2020) Multi-modal classification for human breast cancer prognosis prediction: proposal of deep-learning based stacked ensemble model. IEEE/ACM Trans Comput Biol Bioinform 5963:2–11. https://doi.org/10.1109/TCBB.2020.3018467
109. Sanyal R, Kar D, Sarkar R, Member S (2021) Carcinoma type classification from high-resolution breast microscopy images using a hybrid ensemble of deep convolutional features and gradient boosting trees classifiers. IEEE/ACM Trans Comput Biol Bioinform 5963:1–14. https://doi.org/10.1109/TCBB.2021.3071022
110. Wang D, Khosla A, Gargeya R, Irshad H, Beck AH (2016) Deep learning for identifying metastatic breast cancer, pp 1–6. http://arxiv.org/abs/1606.05718
111. Reis S et al (2017) Automated classification of breast cancer stroma maturity from histological images. IEEE Trans Biomed Eng 64(10):2344–2352. https://doi.org/10.1109/TBME.2017.2665602
112. Feng X et al (2019) Accurate prediction of neoadjuvant chemotherapy pathological complete remission (PCR) for the four subtypes of breast cancer. IEEE Access 7:134697–134706. https://doi.org/10.1109/ACCESS.2019.2941543
113. Siegel RL, Miller KD, Jemal A (2019) Cancer statistics, 2019. CA Cancer J Clin 69(1):7–34. https://doi.org/10.3322/caac.21551
114. Hajela P, Pawar AV, Ahirrao S (2018) Deep learning for cancer cell detection and segmentation: a survey. In: 1st International conference on data science and analytics (PuneCon 2018)—proceedings, pp 1–6. https://doi.org/10.1109/PUNECON.2018.8745420
115. Quellec G, Lamard M, Cozic M, Coatrieux G, Cazuguel G (2016) Multiple-instance learning for anomaly detection in digital mammography. IEEE Trans Med Imaging 35(7):1604–1614. https://doi.org/10.1109/TMI.2016.2521442
116. Sehgal CM, Weinstein SP, Arger PH, Conant EF (2006) A review of breast ultrasound. J Mammary Gland Biol Neoplasia 11(2):113–123. https://doi.org/10.1007/s10911-006-9018-0
117. Youk JH, Gweon HM, Son EJ (2017) Shear-wave elastography in breast ultrasonography: the state of the art. Ultrasonography 36(4):300–309. https://doi.org/10.14366/usg.17024
118. Roslidar R et al (2020) A review on recent progress in thermal imaging and deep learning approaches for breast cancer detection. IEEE Access 8:116176–116194. https://doi.org/10.1109/ACCESS.2020.3004056
119. Hooker KA (2014) Installing concrete anchors. Concr Constr World Concr 59(3):25–30
120. Baker MJ et al (2014) Using Fourier transform IR spectroscopy to analyze biological materials. Nat Protoc 9(8):1771–1791. https://doi.org/10.1038/nprot.2014.110

121. Gonzalez-Hernandez JL, Recinella AN, Kandlikar SG, Dabydeen D, Medeiros L, Phatak P (2019) Technology, application and potential of dynamic breast thermography for the detection of breast cancer. Int J Heat Mass Transf 131:558–573. https://doi.org/10.1016/j.ijheatmasstransfer.2018.11.089
122. Meaney PM, Kaufman PA, Muffly LS, Click M, Poplack SP, Wells WA, Schwartz GN, di Florio-Alexander RM, Tosteson TD, Li Z, Geimer SD (2013) Microwave imaging for neoadjuvant chemotherapy monitoring: initial clinical experience. Breast Cancer Res 15(2):1–16

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

Materials Research Express

PAPER • OPEN ACCESS

An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete

To cite this article: Arti Chouksey *et al* 2022 *Mater. Res. Express* **9** 055003

View the article online for updates and enhancements.

You may also like

- A study based upon the effect of polyvinyl alcohol fiber, iron slag and fly ash over the strength aspects of geo-polymer concrete
farhan Maqbool and Jagdish Chand
- Performance of Alkali-Activated Aerated Concrete using Class C Fly Ash
E Muthu Kumar
- Experimental study of low carbon emission alternative concrete
Muddana Surya Prasanth and Raghava Karumudi



IOP ebooks™

Bringing together innovative digital publishing with leading authors from the global scientific community.

Start exploring the collection–download the first chapter of every title for free.

This content was downloaded from IP address 14.139.251.98 on 27/05/2022 at 11:04

Materials Research Express

PAPER

An investigation on the effect of curing conditions on the mechanical and microstructural properties of the geopolymer concrete

Arti Chouksey[1], **Manvendra Verma**[2,*], **Nirendra Dev**[3], **Ibadur Rahman**[4] and **Kamal Upreti**[5]

[1] Department of Civil Engineering, Deenbandhu Chhotu Ram University of Science and Technology, Murthal, Haryana, India
[2] Research Scholar, Department of Civil Engineering, Delhi Technological University, Delhi, India
[3] Department of Civil Engineering, Delhi Technological University, Delhi, India
[4] Department of Civil Engineering, Jamia Millia Islamia, New Delhi, India
[5] Department of Computer Science and Engineering, Dr Akhilesh Das Gupta Institute of Technology and Management, New Delhi, India
[*] Author to whom any correspondence should be addressed.

E-mail: Arti.civil@dcrustm.org, mv075415@gmail.com, nirendradev@dce.ac.in, Irahman1@jmi.ac.in and Kamalupreti1989@gmail.com

Keywords: geopolymer concrete, curing conditions, oven-curing, mechanical properties, microstructural properties

OPEN ACCESS

RECEIVED
7 March 2022

REVISED
29 March 2022

ACCEPTED FOR PUBLICATION
29 April 2022

PUBLISHED
10 May 2022

Original content from this work may be used under the terms of the Creative Commons Attribution 4.0 licence.

Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

Abstract

Geopolymer concrete represents the future of green and sustainable concrete. It has a large impact on the construction industry owing to its better performance than that of conventional Portland cement concrete. This study aimed to identify the effect of curing conditions on the physical, mechanical, and microstructural properties of specimens using ambient curing and oven-curing. In the experimental analysis, we tested slump and setting time for physical properties, density and drying shrinkage for chemical properties, compressive strength, indirect tensile strength, modulus of rupture, Poisson's ratio, and elastic modulus for mechanical properties, rebound strength, and UPVT for nondestructive and x-ray diffraction, and thermogravimetric analysis for microstructural analysis. After the experimental analysis, it was concluded that the density, Poisson's ratio, and dry shrinkage were higher for ambient-cured specimens than for oven-cured specimens, whereas the compressive strength, indirect tensile strength, modulus of rupture, and elastic modulus of oven-cured specimens were higher than those of ambient-cured specimens. The nondestructive tests, rebound tests, and UPVT show that the oven-cured specimens are better in quality and strength than the ambient cured specimens. In microstructural analysis, x-ray diffraction showed that the oven-cured specimens had a lower intensity of mineral oxides than the ambient-cured specimens in microstructural analysis. The matrix of the ambient-cured specimens was thermally stable up to 800 °C and retained 92% of its original mass, whereas the matrix of the oven-cured specimens retained 94% of its mass up to 800 °C in the thermogravimetric analysis.

Abbreviations

GPC	Geopolymer concrete
OPC	Ordinary Portland cement
GGBFS	Ground granulated blast furnace slag
XRD	X-ray diffraction
TGA	Thermogravimetric analysis
SEM	Scanning electron microscope
EDS	Energy-dispersive x-ray spectroscopy
NaOH	Sodium hydroxide
Na_2SiO_3	Sodium silicate

© 2022 The Author(s). Published by IOP Publishing Ltd

ASTM	American society for testing and materials
ACI	American concrete institute
UPV	Ultrasonic pulse velocity
ELASTIC MODULUS	Modulus of elasticity
SNF	Sulphonated Naphthalene Formaldehyde
PCE	Polycarboxylate ester

1. Introduction

GPC is in high demand in the present era due to its sustainability and economic quality. It is highly resistant to severe conditions for a long duration without degradation of strength. The flyash, GGBFS, metakaolin, rice husk ash, and pozzolanic materials contain high silica & alumina, except for silica fume used as a binding material in the GPC mix design (Verma and Dev 2018). The alkaline solution activates the pozzolans in the GPC mix, containing sodium or potassium hydroxide and sodium or potassium silicates (Krivenko and Kovalchuk 2007). It directly reduces the carbon footprint due to the replacement of the cement content in the concrete mix. Around 8% of carbon dioxide is emitted during the cement production process all over the world. The geopolymerisation reaction occurs during the setting and strengthening of the geopolymer matrix. It contains the various phases of the reaction. The geopolymer matrix is primarily amorphous, and it has a three-dimensional inorganic structure. Equation (1) shows the initial reaction that occurs during geopolymerisation, called dissolution (Xu and Deventer 2002).

After the initial dissolution, the reaction reduces the water to build the end product, shown in the equation (2). Davidovits, in 1978, first gave the name of the reaction and product name called geopolymerisation and geopolymer an inorganic matrix. Geopolymers have different bonding than cement bonding in the matrix during strengthening(Davidovits 1989, Davidovits and Quentin 1991). The geopolymerisation reaction is affected by various parameters, like alkali metals present, silica/alumina ratio, the molarity of sodium or potassium hydroxide, and water content (Verma and Dev 2017). The PCE-based superplasticizer is used to obtain self-compacting concrete in Portland cement concrete (Verma and Nigam 2017), whereas the GPC obtains self-compacting concrete using an SNF-based superplasticizer without any destruction of strength or performance (Verma and Dev 2021c). The liquid-to-binder reaction is critical in the matrix-strengthening reaction (Verma and Dev 2021b). The GGBFS dosage increment in the mix reduces the workability and setting time (Verma and Dev 2021a). The sodium hydroxide molarity and alkaline ratio in the GPC mix design play an essential role in the strength development of concrete (Verma and Dev 2020).

$$\begin{aligned} &n(Si_2O_5,\ Al_2O_3) + \ 2nSiO_2 + \ 4nH_2O + NaOH \text{ or } KOH \downarrow Na^+,\ K^+\ n(OH)_3 \\ &- Si - O - Al^-(OH)_2 - O - S \ - (OH)_3 \end{aligned} \tag{1}$$

$$\begin{aligned} &n(OH)_3 - Si - O - Al^-(OH)_2 - \ O - Si - (OH)_3\overrightarrow{NaOH \text{ or } KOH}\ (Na^+,\ K^+) \\ &(-Si - O - Al^- - O - Si - O -) + nH_2O \end{aligned} \tag{2}$$

Fog-cured samples have a more significant proportion of absorption. This phenomenon might be connected to the geopolymer matrix's ability to retain water, resulting in a more open microstructure in this form of concrete (Bondar *et al* 2010, Bondar *et al* 2015). Heat curing helps speed up the geopolymerisation procedure since compressive strength may be created at a young age (Albitar *et al.* 2015). The oven-drying curing specimens generated 90% of the 28-day compressive strength in 3 days, whereas the ambient-curing specimens attained 57–82 percent. Nevertheless, the ultimate strength of the ambient-curing specimens was somewhat more remarkable than the matching oven-drying cured specimens at 28 days. After 7 days, there is no discernible improvement in compressive strength in the oven-cured specimens (Islam *et al* 2015).

The curing temperatures have an impact on the elastic modulus of alkali-activated concrete. The static modulus of elasticity rose with increasing curing temperature until it reached a limit that appears to be connected to the water to binder ratio at an early age. The static modulus of elasticity falls if water is lost owing to evaporation during curing at a higher temperature before the full strength is achieved. The compressive strength of geopolymer concrete is unaffected by a 60-min rest interval between the casting of specimens and the start of curing (Hardjito *et al* 2004).

The presence of water in the geopolymer and its subsequent removal by evaporation certainly plays a role in achieving a crack-free geopolymer. The fine structure of the microporosity would determine the capillary pressure of occluded water that would promote cracking, and further precise studies in this direction are necessary (Finnie *et al* 2007). The addition of water, 10M NaOH solution, and NSP also affect the flow and strength. The addition of these liquids improves the geopolymer mortars' workability. Water is shown to be the

most effective, resulting in a significant increase in flow with just a minor drop in strength. The addition of a 10 M NaOH solution enhances the flow marginally while keeping the geopolymer mortar's strength. Water and Na ions are necessary for the geopolymerisation process since they are part of the geopolymer system. The addition of NSP to geopolymer mortars does not influence the flow of the mortars (Sathonsaowaphak *et al* 2009). The workability of concrete improves as the amount of time spent mixing it by hand grows, with a maximum of 20 to 30 min being advised (Kumar *et al.*, 2015).

With initial and ultimate setting durations of almost 3 and 10 h, respectively, increasing the mixing time can considerably delay the setting of water glass-activated slag pastes. It provides a practical, effective method of limiting the quick unfavorable setting of these materials, potentially simplifying a wide range of technological applications (Palacios *et al* 2008). When high molarity NaOH is employed, a naphthalene sulphonate polymer-based superplasticizer has little to no influence on the slump and negatively influences fly ash-based geopolymer concrete. When added as an addition to the binder, the conventional superplasticizer usually employed with ordinary Portland cement does not appreciably improve the overall workability of the mix. On the contrary, it reduces the geopolymer matrix's strength. The performance of concrete with superplasticizer at elevated temperatures is similarly lacking. The use of superplasticizers in geopolymer concrete for high-temperature performance is not recommended (Kong and Sanjayan 2010).

The cure period ranged from 4 to 96 h (4 days). The polymerization process was enhanced with a more extended curing period, resulting in increased compressive strength. Up to 24 h of curing time, the rate of development in strength was high; beyond that, the growth in strength was only modest. As a result, the heat-curing duration does not need to exceed 24 h (Lloyd and Rangan 2010). Higher compressive strength is achieved by heating over 6 to 24 h. The rise in strength after 12 h, however, is not considerable (Patankar *et al* 2014).

The longer pre-curing at room temperature in an atmosphere of more than 95% relative humidity is helpful for more significant strength development before heat application (Kani and Allahverdi 2009). The temperature influence depends on the period of curing. Curating at high temperatures for merely 1 h was not responsible for significant changes in strength development, but for prolonged treatment, the reaction rate had been significantly accelerated, and strength had increased at an early age (Rovnaník 2010). The combination of interaction between fly ash and GGBFS controls geopolymerisation at 60 °C. The C-S–H and A–S–H gels in the reaction products confirm this interaction. This enhancement in the compressive strength of the slag addition can be caused by the development and compactness of the microstructure of the gel (C–S–H and A–S-H). Mixing the concrete materials is essential to getting a dense mix or proper mix in the final concrete pouring or molding. The GPC specimens are highly resistant severe conditions compared to conventional concrete specimens (Kumar *et al* 2021, Kumar *et al* 2022).

According to the findings of (Nematollahi and Sanjayan 2014), fly ash based Geopolymer paste activated by only NaOH solution (8.0 M concentration), naphthalene (N) based SP (second generation) was a useful type, resulting in a 136% increase in relative slump without having any adverse effect on the compressive strength when compared to the fly ash-based Geopolymer paste without using any SP. The most effective type of modified polycarboxylate (PC) based SP (latest generation) was found to be the most effective in the case of fly ash based Geopolymer paste activated by multi-compound activator ($Na_2SiO_3/NaOH = 2.5$), resulting in a 39–45 percent increase in relative slump and a maximum 29 percent decrease in compressive strength when compared to the fly ash based geopolymer paste without the use of any SP. It was discovered that the polycarboxylate-based superplasticizer had a retarding impact on alkali-activated fly ash/ slag pastes while having no influence on the heat of hydration, and that it enhanced the workability more considerably than the naphthalene-based superplasticizer in this study. The presence of polycarboxylate-based superplasticizers at concentrations more than 2 weight percent had a favorable influence on the development of compressive strength before 7 days, but it might have a negative effect beyond this time period (Jang *et al* 2014). Because of its intrinsic stability in alkaline media, it is possible to achieve the largest reduction in yield stress in alkali-activated slag cements by including a naphthalene-derivative HRWRA into NaOH-activated slag pastes and mortars. In contrast, the other admixtures used in alkali-activated slag pastes and mortars have little effect on the rheological characteristics of the mixes under consideration (Palacios *et al* 2008).

2. Experimental program

The three major sections are materials and their properties, synthesis or mixing and casting, and experimental test setups.

2.1. Materials

It includes coarse aggregates with a maximum size of 20 mm and M-sand as a fine aggregate on substitute sand, locally accessible primary substances used in the GPC. For triggering the pozzolanic binder, water and

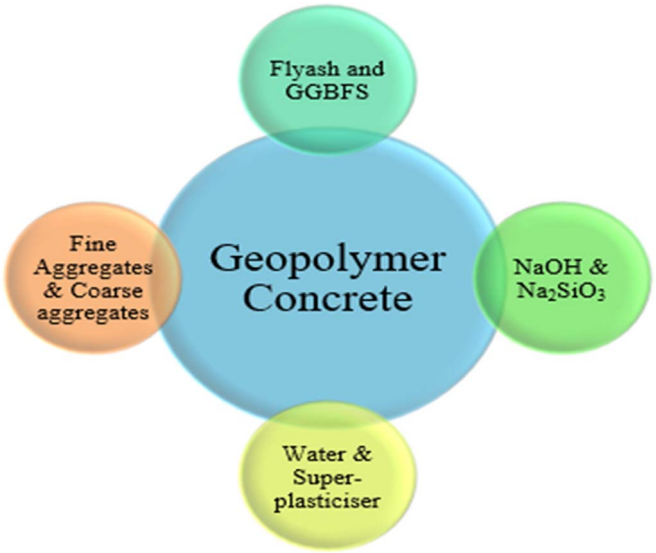


Figure 1. GPC combination of raw materials.

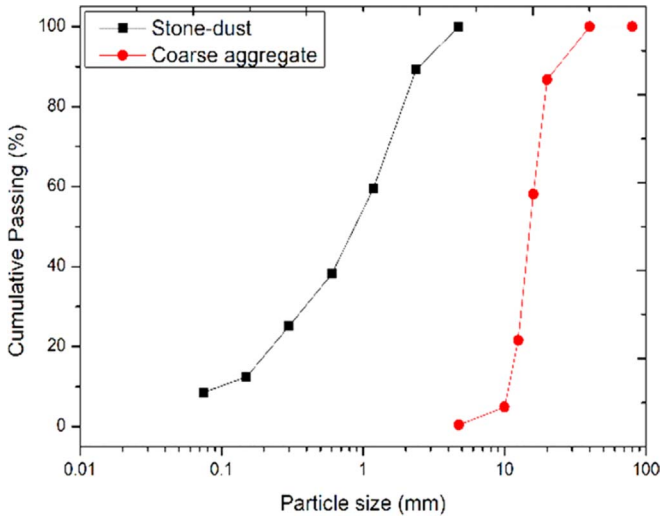


Figure 2. Gradation curve of aggregates.

superplasticizers are used. Flyash and GGBFS were utilized as binders in GPC and alkaline solutions. Figure 1 describes the raw materials in pictorial form to work on a new mix without destroying its strength and durability.

2.1.1. Aggregates

Delhi stone dust sand is used in the laboratory for sampling and testing. The sample weight obtained is 1000g for sand/stone-dust gradation, and the weight retained or passed through many sieves is calculated. Figure 2 shows the gradation curve of the coarse and fine aggregates. The modulus of fineness is an empirical fact that is added to each standard sieve, ranging from 80 mm to 150 mm, and divided by an arbitrary number of 100.That is why the fineness modulus $= (10.7 + 40.5 + 61.9 + 74.9 + 87.6)/100 = 2.756$. The sand is medium since the fineness module ranges from 2.6 to 2.9. The sand is grade II, with a passing percentage of a 600 sieve of between 35 and 59. From the grading curve, the coefficient of 12 is greater than 6, and the coefficient of curvature is 1.33 and is calculated as $D_{10} = 0.1, D_{30} = 0.4, D_{60} = 1.2$. The sand is so well classified. The specific gravity of the substances examined and their values of 2.62 and 1.21%, respectively, for the pycnometer. The silt and bulk densities are 6% and 1610 kg m^{-3}, respectively. Table 1 depicts the properties of fine aggregates (stone dust).

The aggregates 20 mm and 10 mm are employed in the GPC mix design as 60/40 mass ratio. The sieve analysis is conducted for particle size and coarser aggregate graduation. The sample fineness modulus $(13.2 + 41.85 + 78.45 + 95.15 + 99.6 + 5 \times 100)/100 = 7.29$. The fineness modulus, specific gravity,

Table 1. Properties of fine aggregate/stone dust (M-Sand).

S.No.	Test	Results
1.	Zone	Zone II
2.	Grade	Well Graded
3.	Fineness Modulus	2.756 (Medium sand)
4.	Specific Gravity	2.62
5.	Water absorption	1.21%
6.	Silt Content	6%
7.	Bulk density	1610 kg /m^{-3}

Table 2. Properties of coarse aggregate.

S. No.	Test	Results
1.	Fineness Modulus	7.29
2.	Specific Gravity	2.79
3.	Water absorption	0.2%
4.	Crushing Value	23%
5.	Impact Value	22%
6.	Flakiness Index	24%
7.	Elongation Index	30%
8.	Abrasion value	8%

Table 3. Chemical composition.

Minerals	Flyash	GGBFS
Silica (SiO_2)	45.8	34.52
Alumina (Al_2O_3)	21.4	20.66
Lime (CaO)	13.7	32.43
Iron Oxide (Fe_2O_3)	12.6	.57
Magnesia (MgO)	1.3	10.09
Sulphate (SO_3)	1.9	0.77
LOI[a]	0.1	0.3

LOI
[a] Loss of Ignition

water absorption, crushing value, impact value, abrasion value, flakiness index, and elongation index of coarse aggregates were examined in the laboratory. All qualities of coarse units are suggested under the Indian standard regulations for construction purposes. The specific gravity, water absorption, crushing value, impact value, abrasion value, flakiness index, and elongation index of the aggregate samples are 2.79, 0.2%, 23%, 22%, 8%, 24%, and 30%, respectively (IS 2386 Part I (1997), IS 2386 Part III (1997), IS 2386 Part IV (1997), IS 2386 PartV (1997), IS 383 1970, 1997). Table 2 depicts the properties of coarse aggregates.

2.1.2. Alkaline solution

As chemical solution components, sodium hydroxide and sodium silicate activate pozzolanic materials such as flyash and GGBFS. The sodium hydroxide and sodium silicate were purchased from Central Drug House (P) Ltd. The 96 percent minimum test in acidimetric form demonstrates its appearance as a white deliquescent flake with solubility in a 10% solution in water. In the design of the GPC mix, sodium silicate, which is also a component of the alkaline solution, is employed to activate the Flyash and GGBFS. It is a thick, tacky/slightly hazy liquid that is miscible with water and has a minimum assay of Na_2O titrimetric of 10% and a specific gravity of 1.42.

The alkali activators (include sodium hydroxide and sodium silicate) were mixed before 20–24 h, it not includes the superplasticiser. The alkali solution is made before 24 h due the activation of sodium silicate to react with the binding materials mineral oxides.

2.1.3. Superplasticizer and water

Conplast SP 430 is based on SNF, utilized as a superplasticizer in the GPC design mix. It has the appearance of a dark liquid and has up to 2% air entrainment with a specific gravity of 1.18 at 25 °C (IS 9103 1999 1999).

Tape water is used in the GPC mix design, which has a 6-pH level.

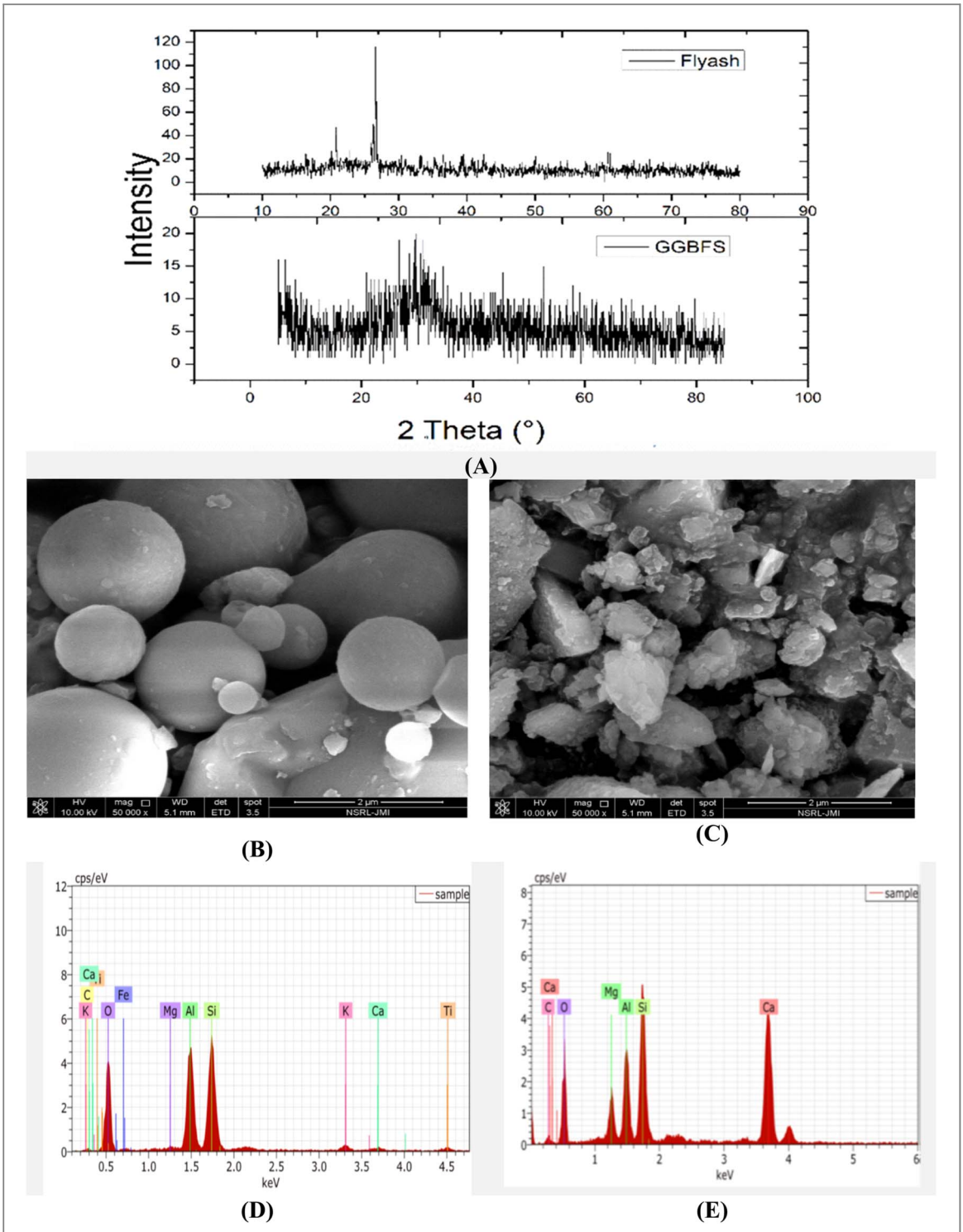


Figure 3. (a) XRD graph of flyash and GGBFS; (b) SEM image of flyash; (c) SEM image of GGBFS; (d) EDS Graph of Flyash; (e) EDS Graph of GGBFS.

2.1.4. Pozzolans

Flyash is produced in thermal power plants as a byproduct of coal combustion for electricity generation. Flyash is harvested at the Guru Nanak Dev Thermal Plant in Bhatinda, Punjab, India. Flyash is classified as Class C in the ASTM C618–33512 standard (ASTM C 618 19, 2019). Table 3 provides the chemical compositions of the sample materials, and figure 3(b) shows a scanning electron microscopy image of the porous spherical flyash particles. The elemental component graph of the material analyzed by EDS is depicted in figure 3(d). The XRD graph of samples is shown in the figure 3(a).

GGBFS is produced by isolating molten slag with the assistance of high-pressure water jets. Because molten slag separation precludes crystallization, the final products are glassy and granular aggregates. The finished



Figure 4 (a) Picture of Pan mixture during mixing; (b) Picture of oven curing of cube specimens; (c) Picture of ambient-curing of cube specimens.

product is crushed, pulverized, and segregated for a variety of purposes. GGBFS is working on a project at the Bhilai Steel Plant in Durg, Chhattisgarh, India. Table 3 reveals the chemical contents of the sample materials, and figure 3(c) shows a scanning electron microscopy image showing the porous indeterminate form of GGBFS particles. The elemental component graph of the material analyzed by EDS is depicted in figure 3(e).

2.2. Synthesis

For one batch, all of the basic materials are mixed into a 40-litre capacity mix. Figure 4(a) shows a picture of the pan mixture during mixing. The alkaline chemical and the sodium hydroxide, sodium silicate, and superplasticizer were mixed 24 h before casting. At the time of mixing, pour the coarse aggregate into the combination, then the sand or fine aggregate and thoroughly mix, then the chemical into the mix, and finally the extra water as needed during the mixing. Mix the concrete for 10–15 min in the pan mixture (IS 12119 1987, 1999). The appropriately mixed concrete plastic paste is formed in the moulds for the cubes, cylinders, and beams (IS 1199 1959, 1959, IS 10086 1982, 2008). After three days of casting, the specimens were de-molded and cured in the oven for 24 h, shown in the figure 4(b), and in ambient-curing, shown in figure 4(c). Table 4 summarizes the raw material mix composition for the GPC. The molarity of the sodium hydroxide and liquid to binder ratio are taken 14 M, .6 from the references (Verma and Dev 2021b).

There are two types of curing is showed in the manuscript that are ambient-curing and oven-curing. In the ambient curing, the samples were kept in the open surroundings or environment without any changes. The samples were hardened and strengthen continuously with the increment of age. In the oven-curing, the samples were kept in the oven for 24 h at temperature of 80 °C after demoulding. The humidity factor is not included in the research parameters, but it would be maintained around 90–95.

Table 4. Mix design of GPC.

Material	Quantity (kg m)$^{-3}$
Flyash	303.75
GGBFS	101.25
Coarse Aggregates (20 mm)	761
Coarse Aggregates (10 mm)	508
Fine Aggregates/Stone dust	683
Sodium Hydroxide	46.28
Sodium silicate	115.72
Superplasticiser	4.05
Water	20.25

2.3. Test setups

All experiments are carried out in the concrete laboratory on the GPC mix in its fresh or hardened state.

2.3.1. Physical properties tests

The slump test is used to determine the workability of a fresh geopolymer concrete blend. The top diameter, bottom diameter, and height of the slump apparatus are 100 mm, 200 mm, and 300 mm, respectively. Millimetres are used to measure the slump (IS 7320 1974, 2008).

The setting time of the fresh GPC was found through the penetrometer test on the concrete sample. It determines both the initial setting time and the final setting time of the fresh sample (ASTM C403/C403M 16 2019).

2.3.2. Chemical properties tests

Density is used to assess the chemical properties of GPC mix samples. Before the 28-day destructive inspection, the weight of cube samples was used to test the density of the mix design. It directly indicates the hardness and strength of the concrete specimens (ASTM C138/C138M 17a 2019).

The drying shrinkage occurs in the concrete specimens after the evaporation of water from the mix or formation of end products of the bond because end products take less volume than the volume of the raw products, and the length comparator of the concrete or mortar specimens is at the micron reading level (ASTM C157/C157M 08 2008).

2.3.3. Mechanical properties tests

The GPC mix designs cubic samples of dimension 150 mm *150 mm *150 mm are used to test the compressive strength of the specimens. The samples were tested in an axial loading mode on a universal Testing Machine at a loading rate of 5.2 kN/s (IS 516 1959, 2004).

The splitting tensile strength of cylindrical samples with diameters and heights of 150 mm and 300 mm is measured. In the universal testing machine, transverse loads are applied to the cylinder to investigate the splitting tensile strength of the GPC cylinder specimens (IS 5816 1999 1999).

The width, height, and length of the beam sample are 100 mm, 100 mm, and 500 mm, respectively, for measuring the flexural strength of the GPC mix design. At the flexural testing machine, the beam specimen was subjected to a two-point load or flexural tensile test to determine the flexural strength of the GPC mix specimen (IS 9399 1979, 1979).

The cylindrical samples have a diameter of 150 mm and a length of 300 mm, and they would be used to measure the modulus of elasticity and Poisson's ratio of the GPC mix design. In the universal testing machine, the axial load is applied to the cylinder to analyze the vertical and horizontal strain of the GPC cylinder specimen to measure the modulus of elasticity and Poisson's ratio(ASTM C469/C469M 14 2019).

2.3.4. Non-destructive tests

Those tests conducted in the laboratory and the field to detect strength or quality without destruction are non-destructive tests.

The rebound test is based on the apparatus's check of surface hardness, and it is performed on cube samples 7 days, 14 days, 28 days, 42 days, and 56 days after casting. It was performed on all types of sample cubes and cylinders to determine the strength of the mixed sample (IS 13311-2 (1992)).

The ultrasonic pulse wave frequency travels through the sample in the UPVT. The greater the UPV indicates, the greater the strength and efficiency of the GPC specimen. Two transducers, an electrical pulse generator, an amplifier, and an electronic timing unit, are used in the research apparatus. The ultrasonic pulse wave is sent through the transducers and an electronic timing machine (IS 13311 Part1 1992).

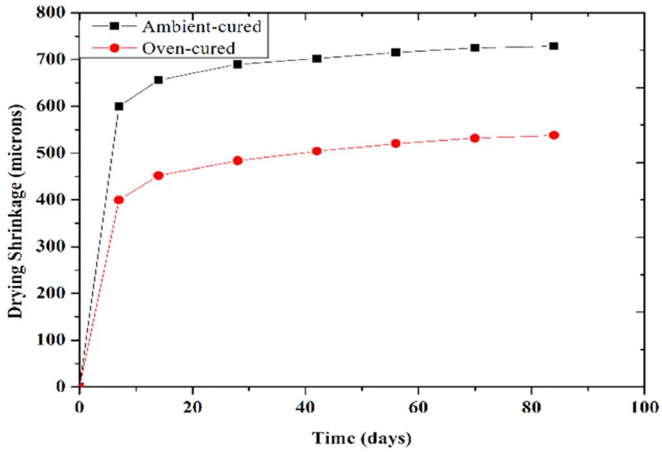


Figure 5. Drying shrinkage variation of both cured specimens.

2.3.5. Microstructural analysis

The microstructural analysis checks the effect of various parameters on the geopolymer matrix chemistry.

The bond-forming or chemical reaction end products phase analysis is performed in the laboratory using an XRD test to determine if the phase formation is crystalline or amorphous with oxides (ASTM D934 13, 2014).

The TGA tests the thermal stability of the materials up to 800 °C. The temperature increases at 10 °C m^{-1} and continuously checks the sample's weight retention from room temperature to 800 °C.

3. Results and discussion

All the experimental tests and their results are described briefly in this section and discussed in the prior works of literature.

3.1. Physical properties

The physical property analysis tests the workability and initial & final setting time of the GPC. The slump test is used to find workability, whereas the penetrometer test is used to find the initial and final setting time of the fresh GPC. Table 4 describes the mixed design of the raw content of the GPC. The value of the fresh GPC is found at 75 mm, whereas the initial setting time and final setting time are 70 min and 160 min, respectively.

3.2. Chemical properties

Their chemical properties mean that chemistry directly plays a vital role in their denseness and their shrinkage. It includes the density and drying shrinkage of the GPC mix design specimens.

3.2.1. Density

The density of the GPC mix design was determined by the cube specimen's weight 28 days after its casting. It is easily calculated through the equation of mass by volume. The ambient-cured specimens' density is higher than the oven-cured specimens. The density of the ambient cured specimens, and oven-cured specimens average 2494 kg m^{-3} and 2487 kg m^{-3}. Hardjito *et al* (2004) also stated that the density of the ambient-cured flyash-based GPC is higher than the oven or heat cured specimens.

3.2.2. Drying shrinkage

The drying shrinkage of the GPC mix design was determined through the length comparator tested at 7, 14, 28, 42, 56, 70, and 84 days after the casting. Figure 5 describes the graph of both cured GPC mix specimens tested on various days. It shows that the ambient cured GPC mix specimens got higher drying shrinkage than the oven-cured specimens. The dried shrinkage occurred in 95% of the tests, with only minor variations. The ambient cured specimen and oven-cured specimen maximum drying shrinkage found in the 84-day test was 729 microns and 538 microns. According to (Palacios and Puertas 2011), the drying shrinkage of ambient-cured specimens is greater than that of heat-cured specimens (Palacios *et al* 2008, Palacios and Puertas 2011). Drying shrinkage occurs in the conventional concrete due to the evaporation of capillary water from matrix, and its increases with the increment of surrounding temperatures, but in the geopolymer concrete the wetting curing is not done. The oven-cured specimens fasten the reaction and make the bonds faster and retain the water content in the matrix,

whereas in the ambient-cured specimens' matrix the reaction occurring is slower and the water evaporation are higher with the time.

3.3. Mechanical properties

The mechanical properties of the GPC mixed design are determined through the various tests being conducted on the specimens. It includes the poison ratio, elastic modulus, compressive strength, indirect tensile strength, and flexural strength of the GPC mix design.

3.3.1. Poisson's ratio

The Poisson's ratio of the GPC mix design is determined in the UTM cylinder specimen test. The Poisson's ratio is directly calculated by the ratio of lateral strain to the linear strain. These strains are determined by the testing of the cylinder in UTM applying longitudinal loading on the specimen. LVDT is used to directly get the results of strains, including both linear and lateral. The ambient cured specimens have a higher Poisson's ratio than the oven-cured specimens. Ambient-cured and oven-cured specimens have Poisson's ratios are .16 and.14, respectively.

3.3.2. Elastic modulus

The cylinder specimens determine the elastic modulus of the GPC mix design. It is calculated through the stress-strain relation using a stable modulus as per the IS code. It demonstrates that the elastic modulus of ambient-cured specimens is lower than the elastic modulus of oven-cured specimens. The elastic modulus of ambient cured and oven cured specimens is 22.1GPa and 24GPa, respectively.

3.3.3. Compressive strength

Concrete's compressive strength, also known as its characteristic strength, is a fundamental mechanical property. It is determined through the tests conducted on the cube specimens as per the IS code in the UTM. The uniaxial static load was applied during the compressive strength test on cube specimens, and figure 6(a) describes the graph between compressive strength versus time of both cured specimens. The tests were performed on specimens 7, 14, 28, 42, and 56 days that had been cast on both ambient cured and oven-cured specimens. It shows that the oven-cured specimens have a higher compressive strength than the ambient cured specimens.

The specimen's compressive strength was found to be around 95% after a 28-day test. Nagalia *et al* (2016) stated that the compressive strength increases with the temperature increment (Nagalia *et al* 2016). According to (Heah *et al* 2011), ambient-curing of GPC specimens delayed setting time and resulted in more significant strength formation than ambient-cred specimens (Heah *et al* 2011). The compressive strength of the ambient-cured and oven-cured specimens was 24.8 MPa and 35 MPa after 28 days, respectively.

3.3.4. Splitting tensile

The splitting tensile strength is also called indirect tensile strength, determined through the tests conducted on the cylinder specimens in which the load was applied statically along the lateral direction of the specimens. This test is also conducted on the UTM. Figure 6(b) describes the graph between the splitting tensile strength versus time of both ambient-cured and oven-cured specimens.

It shows that the oven-cured specimens have a higher strength than the ambient-cured specimens. The tests were performed on specimens 7, 14, 28, 42, and 56 that had been cast on both ambient cured and oven-cured specimens. It also has a similar strength pattern to compressive strength. The maximum splitting tensile strength of ambient-cured and oven-cured specimens is 4 MPa and 5.2 MPa at 56 days, respectively.

3.3.5. Flexural strength

The modulus of rupture is another name for flexural strength. It could be determined through the prism shape of specimens tested in the flexural testing machine. All of the specimens were subjected to a two-point static load in the lateral direction. The tests were performed on specimens 7, 14, 28, 42, and 56 days that had been cast on both ambient cured and oven-cured specimens.

Figure 6(c) describes a graph between the flexural strength and age of concrete specimens of both cured samples. It shows a similar pattern of flexural strength to the compressive strength and splitting tensile. The oven-cured specimens have a higher flexural strength than the ambient-cured specimens. A 28-day test of both ambient-cured and oven-cured specimens revealed 98% flexural strength. The flexural strength of the ambient-cured and oven-cured GPC specimens was 3.8 MPa and 5.5 MPa in a 28-day test, respectively. It has a compressive strength of approximately 16%. The ratio of flexural strength to the compressive strength of the GPC specimens is higher than conventional Portland cement concrete (Joseph 2015).

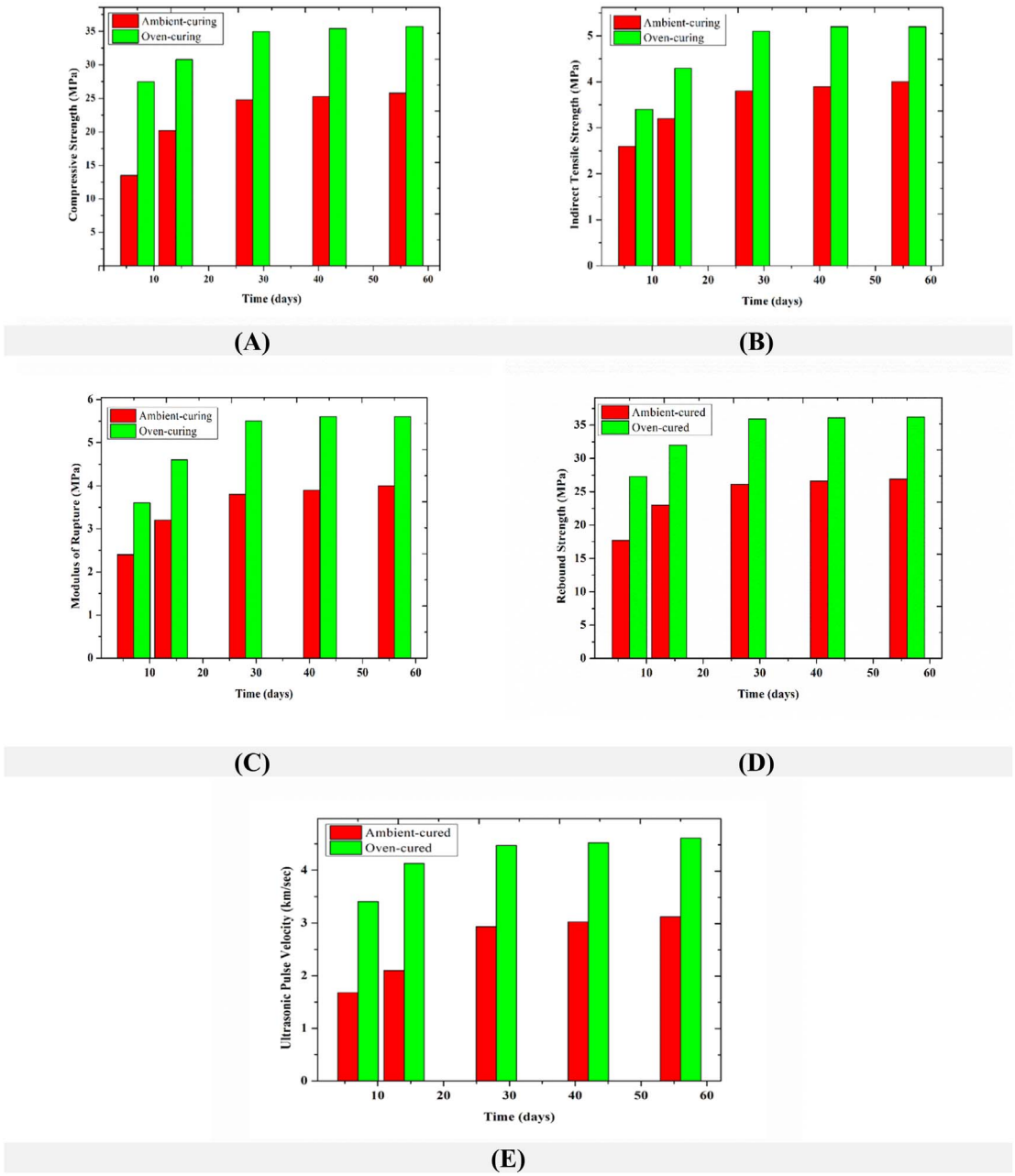


Figure 6. (a) Graph between compressive strength versus time; (b) Graph between the splitting tensile versus time; (c) Graph between the flexural strength versus time; (d) Graph between the rebound strength and age of GPC; (e) Graph between UPV versus age of GPC.

3.4. Non-destructive tests

The non-destructive test is used to determine the strength or quality of a concrete specimen without causing it to be destroyed. It comprises the rebound hammer test as well as the UPV test.

3.4.1. Rebound strength

According to the IS code, the rebound strength shows the similar compressive strength of the concrete specimens with a precision of 20%. The surface hardness of the specimen determines the specimen's strength. All cube specimens of both cured conditions were tested for rebound strength on 7, 14, 28, 42, and 56 days after casting and before the destructive test on the samples. The rebound strength of oven-cured samples is greater than that of ambient-cured samples of the same GPC mix, and figure 6(d) describes the graph of rebound hammer test strength with the age of the GPC of both ambient-cured and oven-cured specimens. It shows a very similar pattern of strength to the compressive strength of the GPC specimens.

The rebound strength of ambient-cured specimens and oven-cured specimens is 26.1 MPa and 35.9 MPa after 28 days, respectively. Thus, in the 28-day tests performed on specimens of both cured conditions, approximately 98% of the strength was measured.

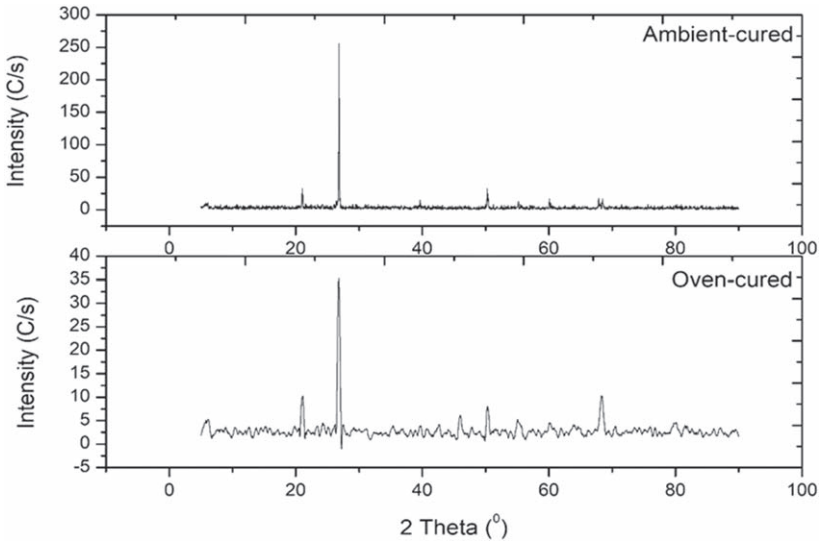


Figure 7. XRD graph of ambient and oven-cured specimens' matrix.

3.4.2. UPVT

The UPV test is used to determine the intensity or quality of an ultrasonic pulse wave by monitoring its passing time. The time would increase if the quality or strength of the specimens deteriorated. The UPV test detects the development of a crack inside the specimen. UPV tests are performed on cube samples of GPC mix 7 days, 14 days, 28 days, 42 days, and 56 days after casting and before destructive testing on the samples. The device measures the duration of the UPV wave and computes the velocity, which determines the quality of GPC samples.

Figure 6(e) describes the graph between UPV versus age of the GPC of both ambient-cured and oven-cured specimens. It shows a similar pattern to the rebound strength graph, which shows that the oven-cured specimen's UPV is higher than the ambient-cured specimens. After 56 days of testing, the maximum UPV of ambient-cured and oven-cured specimens is 3.12 km s^{-1} and 4.61 km s^{-1}, respectively.

3.5. Microstructural analysis

The microstructural analysis of the GPC matrix shows the internal chemistry occurs in the different conditions determined through the various tests: XRD analysis, FTIR analysis, SEM analysis, and TGA-DTG analysis. In addition, it describes the internal chemistry that directly affects the performance of the specimen, which includes both strength and durability properties.

3.5.1. XRD

The XRD examination of the materials defines the intensity of the different oxides contained in the material. The XRD test is performed on tiny particles in powder form that are less than 150 microns in size. It specifies the intensity of the mineral oxide count in the sample. Figure 7 depicts an XRD graph of GPC samples, including (A) ambient-cured GPC specimens and (B) oven-cured GPC specimens. It demonstrates that the intensity of the oven-cured sample is lower than that of the ambient-cured sample due to the decrease of the oxide mineral components during oven-curing. However, the intensity pattern of both cured samples follows a similar trend. After 28 days, a sample is obtained from the destructive test on the cube and chosen as the optimal strength combination for the XRD test. The XRD test results describe the crystal shape of the mineral oxides and their intensities in the correct pattern. The quartz (silica) has the most prominent peak in the pattern in both cured conditions.

3.5.2. SEM Analysis

The SEM analysis of both conditions cured hardened GPC specimens depicts the differences of the chemical reaction that occurred, in which check the image at the 2-micron level. The ambient-cured specimens hardened linearly with the time passage, whereas the oven-cured specimens got the strength or hardened spontaneously. The oven-drying condition accelerates the geopolymerisation reaction and strengthens the specimens at an early age. Figure 8 shows the SEM images of (a) ambient-cured specimen and (b) oven-cured GPC specimen. It shows that the ambient-cured specimens show fewer pores than the oven-cured specimen image. The oven-cured

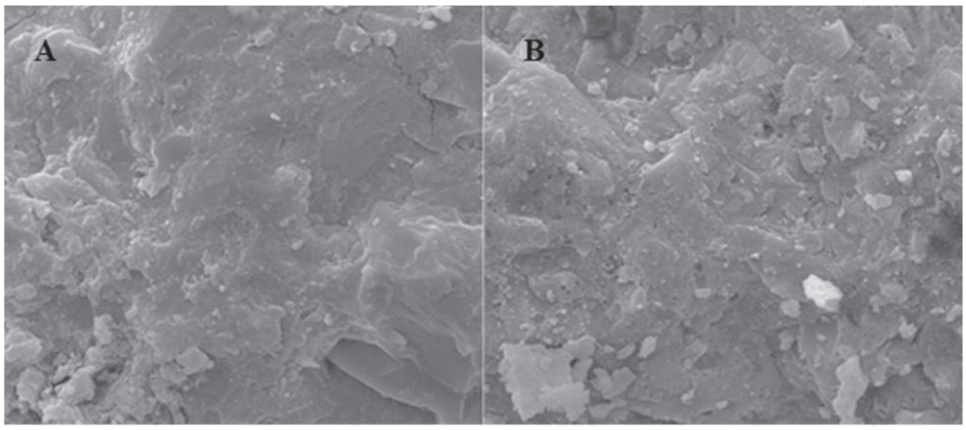


Figure 8. SEM graph of (A) ambient-cured (B) oven-cured GPC.

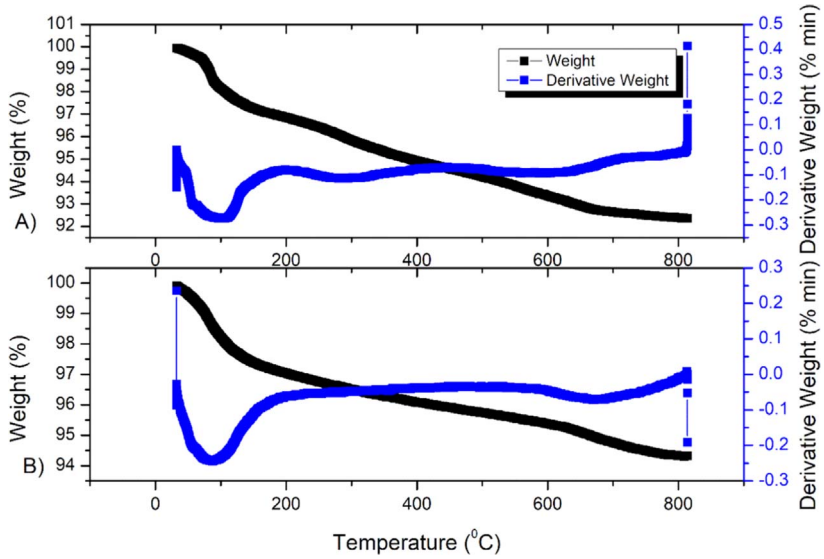


Figure 9. TGA-DTG figure of both cured GPC matrix.

specimens reduce the water from the matrix after curing. The geopolymer paste matrix provides honeycomb structures after the elevated temperature exposure (Fan *et al* 2018).

3.5.3. TGA

TGA-DTG tests provide thermal analysis of materials, evaluating weight as the temperature increases at a rate of 10 °C per minute up to 800 degrees celsius. The TGA-DTG test sample is produced in the lab by grinding and sieving the materials to 150-microns. The graph of TGA-DTG of A) ambient-cured B) oven-cured specimens of the GPC mix is shown in the figure 9. It demonstrates that the weight preserved as a percentage of oven-cured samples is more significant than ambient-cured samples up to 800 °C. The weight preserved by ambient-cured and oven-cured samples is 92% and 94%, respectively, since some water evaporates during the curing time in the oven-cured specimens. In addition, the sample materials have far better stability than other concrete materials, such as OPC concrete. At 800 °C, it lowers weight by up to 8% and exhibits a similar derivative pattern.

4. Correlation between mechanical properties

GPC mechanical characteristics demonstrate a linkage to each other after a long-term experimental examination. Based on the findings of destructive experiment testing of various specimens, the correlation equation between compressive strength, splitting strength, flexural strength, and elastic modulus was presented.

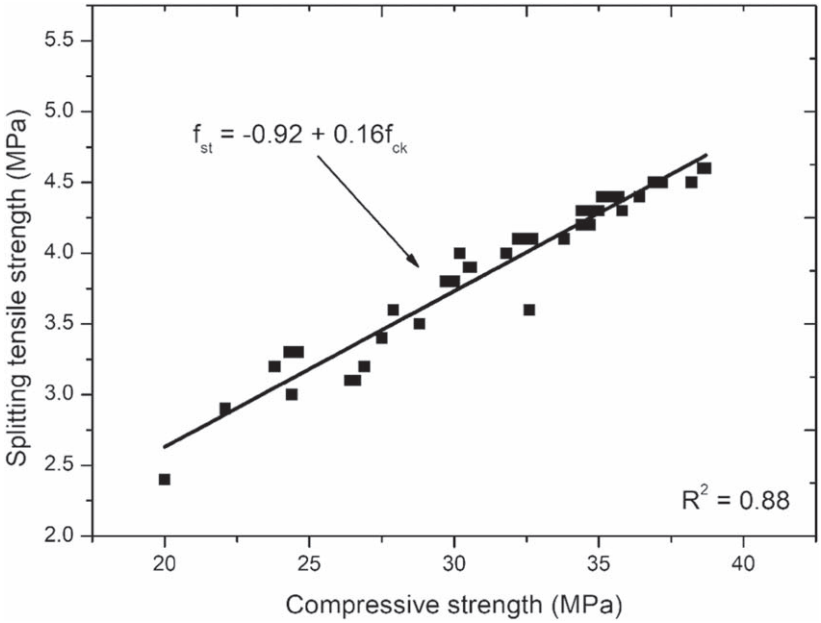


Figure 10. Correlation graph between the compressive strength and splitting tensile strength.

4.1. Correlation between splitting tensile strength and compressive strength

The proposed equation (3) between the compressive strength and splitting strength relationship in fly ash-slag-based GPC was developed through experimental research into various tests performed on the same mix design and using two types of curing: ambient-curing and oven-curing. Figure 10 shows the graph of correlation equation generation between compressive strength and splitting tensile strength. Equation (4) from the ACI committee's report on high strength concrete and equation (5) presented by the Euro-International committee for concrete in CEB/FIP in 1990 (International, C of E 1990, Russell *et al* 1997). The equation (6) for the connection between splitting and compressive strength is provided by an ACI standard and report (ACI 318–14) (Committee 2014). The connection equation was proposed by Ahmed *et al* (7), and Ryu *et al.* suggested the equation (8) to predict splitting strength based on GPC compressive strength(Ryu *et al* 2013, Ahmed *et al* 2020). Gardener *et al* represented the correlation between the splitting tensile equation (9), whereas Raphael *et al* proposed a correlation equation (10) between splitting tensile and compressive strength (Gardner and Poon 1976, Gomes *et al* 2020). In their book published by Tayler & Francis, De-larrard also proposed a correlation equation (11) (Larrard 1999). Table 5 depicts the all-correlation equations described in the literature and analyses the various correlation equations numerically between the splitting tensile and compressive strength.

$$Proposed\quad Eq.: f_{st} = -0.92 + 0.16 f_c \tag{3}$$

$$ACI363R-92: \quad f_{st} = 0.59\sqrt{f_c} \tag{4}$$

$$CEB-FIP: \quad f_{st} = 0.301 f_c^{0.67} \tag{5}$$

$$ACI318-14: \quad f_{st} = 0.56 f_c^{0.5} \tag{6}$$

$$Ahmed\ et\ al: \quad f_{st} = 0.462 f_c^{0.55} \tag{7}$$

$$Ryu\ et\ al: \quad f_{st} = 0.17 f_c^{0.75} \tag{8}$$

$$Gardener\ et\ al: \quad f_{st} = .6 f_c^{2/3} \tag{9}$$

$$Raphael\ et\ al: f_{st} = f_c^{.7} \tag{10}$$

$$De-larrard: f_{st} = .6 + .06 f_c \tag{11}$$

Where; f_c is characteristic strength in MPa, f_{fs} is flexural strength in MPa

4.2. Correlation between flexural strength and compressive strength

In the oven curing situation, GPC specimens of various mixes obtained compressive strength and flexural strength, which provides the equation (12) of the correlation between the compressive strength and flexural strength and the other equations presented by other authors or codes. Figure 11 shows the graph of correlation

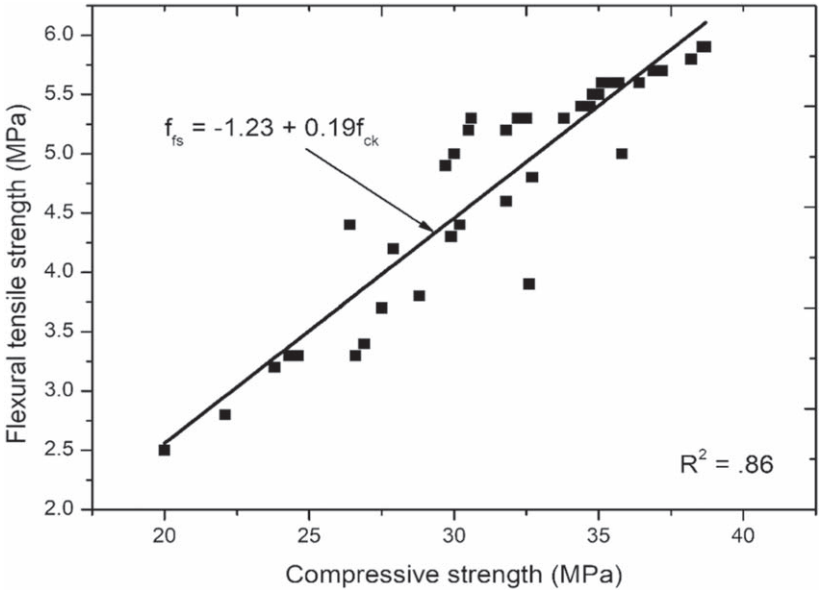


Figure 11. Correlation graph between the compressive strength and flexural tensile strength.

Table 5. Correlation between the splitting tensile and compressive strength equations.

fc	Proposed Eqs. fs = 0.25 (fc)^0.73	ACI 363R-92 fs = .59 (fc)^.5	CEB-FIP fs = .301 (fc)^.67	ACI 318–99 fs = .56 (fc)^.5	Ahmad *et al* fs = .462 (fc)^.55	Ryu *et al* fs = .17 (fc)^(3/4)	Gardener *et al* fs = .6(fc)^ (2/3)	Raphael *et al* fs = .2 (fc)^.7	De-lar-rard *et al* fs = .6 +.06fc
15.00	1.81	2.29	1.85	2.17	2.05	1.30	3.65	1.33	1.50
17.50	2.02	2.47	2.05	2.34	2.23	1.45	4.04	1.48	1.65
20.00	2.23	2.64	2.24	2.50	2.40	1.61	4.42	1.63	1.80
22.50	2.43	2.80	2.42	2.66	2.56	1.76	4.78	1.77	1.95
25.00	2.62	2.95	2.60	2.80	2.71	1.90	5.13	1.90	2.10
27.50	2.81	3.09	2.77	2.94	2.86	2.04	5.47	2.03	2.25
30.00	2.99	3.23	2.94	3.07	3.00	2.18	5.79	2.16	2.40
32.50	3.17	3.36	3.10	3.19	3.13	2.31	6.11	2.29	2.55
35.00	3.35	3.49	3.26	3.31	3.26	2.45	6.42	2.41	2.70
37.50	3.52	3.61	3.41	3.43	3.39	2.58	6.72	2.53	2.85
40.00	3.69	3.73	3.56	3.54	3.51	2.70	7.02	2.65	3.00

equation generation between compressive strength and flexural tensile strength. The ACI Committee's equation (13) in the state-of-the-art report on high-strength concrete and equation (14) are elaborated on in the ACI committee's building code requirements for structural concrete. AS 3600–2009 specifies the RCC design technique and equation (15) relating flexural strength to compressive strength (Concrete 2009). Equation (16) from the Indian code of practice for plain and reinforced concrete and equation (17) from Ahmed's 2014 international journal of structural engineering, vol. 5. Bellum *et al* presented the link between compressive and flexural strength in equation (18). Mhaiskar and Naik proposed a correlation equation (19), whereas Juki *et al* the equation (20) (Juki *et al* 2013). Table 6 depict the suggested equation in conjunction with the other related equations.

$$Proposed \quad Eq.: \quad f_{fs} = -1.23 + 0.19 f_c \tag{12}$$

$$ACI363R - 92: \quad f_{fs} = 0.94\sqrt{f_c} \tag{13}$$

$$ACI \quad 318 - 99: \quad f_{fs} = 0.62\sqrt{f_c} \tag{14}$$

$$AS \quad 3600: \quad f_{fs} = 0.6\sqrt{f_c} \tag{15}$$

$$IS \quad 456 - 2000: f_{fs} = 0.7\sqrt{f_c} \tag{16}$$

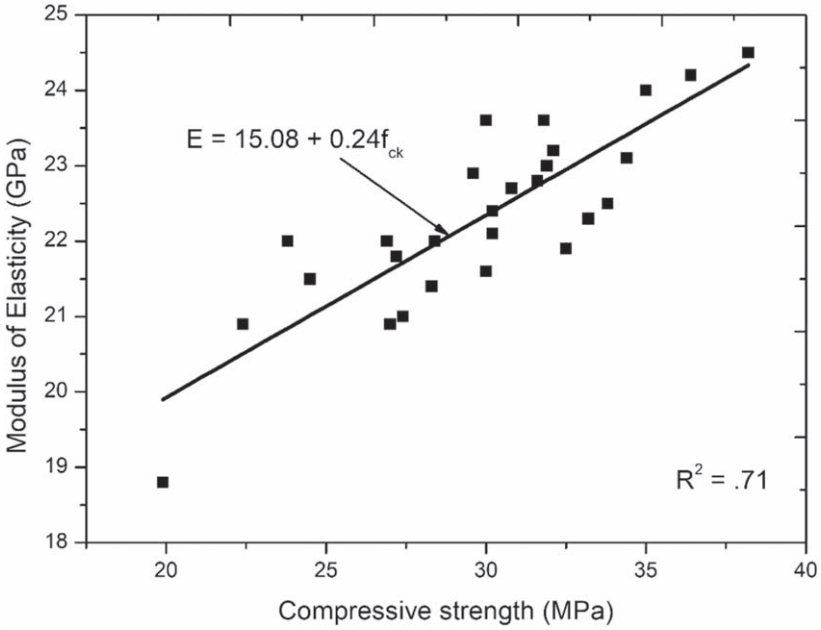


Figure 12. Correlation graph between the compressive strength and elastic modulus.

Table 6. Correlation between the splitting tensile and compressive strength equations.

fc	Proposed equation fs = .25 (fc)^0.85	ACI 363R-92 F = 0.94 (fc)^.5	ACI 318–99 F = 0.62 (fc)^.5	AS 3600 F = 0.6 (fc)^.5	IS 456: 2000 fs = .7 (fc)^.5	M. Ahmed *et al* F = 1.055 (fc)^.5	Bellum *et al* F = 0.57 (fc)^.74	Mhaiskar & Naik F = .864 (fc)^.5	M. Irwan Juki *et al* F = .466 (fc)^.703
15	2.50	3.64	2.40	2.32	2.71	4.09	4.23	3.35	3.13
17.5	2.85	3.93	2.59	2.51	2.93	4.41	4.74	3.61	3.49
20	3.19	4.20	2.77	2.68	3.13	4.72	5.23	3.86	3.83
22.5	3.53	4.46	2.94	2.85	3.32	5.00	5.71	4.10	4.16
25	3.86	4.70	3.10	3.00	3.50	5.28	6.17	4.32	4.48
27.5	4.18	4.93	3.25	3.15	3.67	5.53	6.62	4.53	4.79
30	4.50	5.15	3.40	3.29	3.83	5.78	7.06	4.73	5.09
32.5	4.82	5.36	3.53	3.42	3.99	6.01	7.49	4.93	5.39
35	5.13	5.56	3.67	3.55	4.14	6.24	7.92	5.11	5.67
37.5	5.44	5.76	3.80	3.67	4.29	6.46	8.33	5.29	5.96
40	5.75	5.95	3.92	3.79	4.43	6.67	8.74	5.46	6.23

$$M.\ \ Ahmed: \quad f_{fs} = 1.055\sqrt{f_c} \tag{17}$$

$$Bellum\ \ et\ \ al: \quad f_{fs} = 0.57 f_c^{0.74} \tag{18}$$

$$Mhaiskar\ \&\ Naik\ \ : \quad f_{fs} = .864\sqrt{f_c} \tag{19}$$

$$Juki\ \ et\ \ al: \quad f_{fs} = .466 f_c^{.703} \tag{20}$$

Where; f_c is characteristic strength in MPa, f_{fs} is flexural strength in MPa

4.3. Correlation between the elastic modulus and compressive strength

The elastic modulus assesses the resistance capacity of the elastic deformation caused by the applied load. GPC's elastic modulus was calculated using ASTM C469/469M-14. Following the experimental research, the connection equation (21) between elastic modulus and compressive strength was presented for elastic modulus prediction based on compressive strength. Figure 12 shows the graph of correlation equation generation between compressive strength and elastic modulus. The elastic modulus of typical concrete is anticipated using equation (22) for density ranges ranging from 1442 to 2483 kg m^{-3} (Committee 2014), according to ACI 318–14

Table 7. Correlation between the splitting tensile and compressive strength equations.

fc	Proposed Eq Ec = (15080 + 240fc)	ACI 318–14 Ec = .043*p^(1.5)*(fc)^.5	CEB-FIP Ec = .85*2.15*10000*(fc/10)^(1/3)	IS 456–2000 Ec = 5000(fc)^.5	Lee & Lee Ec = 5300(fc)^(1/3)	Nath & Sarker Ec = 3510*(fc)^.5	Bellum *et al* Ec = 3282*(fc)^.5	Hardjito *et al* Ec = 2707(fc)^.5 + 5300	Diaj-loya *et al* Ec = .037*p^1.5*(fc)^.5
15.00	18680.00	20567.98	20919.65	19364.92	13070.92	13594.17	12711.13	15784.17	17698.03
17.50	19280.00	22215.96	22022.67	20916.50	13760.11	14683.38	13729.59	16624.19	19116.05
20.00	19880.00	23749.85	23025.06	22360.68	14386.41	15697.20	14677.55	17406.07	20435.92
22.50	20480.00	25190.53	23947.02	23717.08	14962.47	16649.39	15567.89	18140.43	21675.57
25.00	21080.00	26553.15	24802.99	25000.00	15497.29	17550.00	16410.00	18835.00	22848.06
27.50	21680.00	27849.17	25603.63	26220.22	15997.55	18406.60	17210.95	19495.63	23963.24
30.00	22280.00	29087.51	26357.11	27386.13	16468.33	19225.06	17976.25	20126.85	25028.79
32.50	22880.00	30275.24	27069.81	28504.39	16913.64	20010.08	18710.28	20732.27	26050.79
35.00	23480.00	31418.10	27746.83	29580.40	17336.65	20765.44	19416.57	21314.83	27034.18
37.50	24080.00	32520.83	28392.34	30618.62	17739.97	21494.27	20098.06	21876.92	27983.04
40.00	24680.00	33587.37	29009.75	31622.78	18125.75	22199.19	20757.19	22420.57	28900.76

(ACI Building Code). For typical concrete, the CEB-FIP provides an equation (23) for elastic modulus prediction based on compressive strength (International, C of E 1990). The Indian standard code represents the equation (24) for estimating elastic modulus for average weight conventional concrete, IS 456 2000 (2000) Lee and Lee suggested equation (25) to predict the elastic modulus of GPC (Lee and Lee 2013) and Nath and Sarker provided equation (26) to predict the elastic modulus of ambient-cured low-calcium fly ash blended GPC; (Nath and Sarker 2017). Bellum *et al.* suggested equation (27) to estimate elastic modulus using fly ash-GGBFS-based GPC (Bellum *et al* 2019). Hardjito *et al.* proposed a correlation equation between elastic modulus and compressive strength of the flyash-based GPC. Table 7 depicts the suggested equation in conjunction with the other related equations.

$$Proposed \quad Eq.: \quad E_c = 15.08 + 0.24 f_c \tag{21}$$

$$ACI318 - 14: \quad E_c = 0.043 \times \rho^{1.5} \times \sqrt{f_c} \tag{22}$$

$$CEB - FIP: \quad E_c = 0.85 \times 2.15 \times 10^4 (f_c/10)^{1/3} \tag{23}$$

$$IS \quad 456 - 2000: \quad E_c = 5000\sqrt{f_c} \tag{24}$$

$$Lee \quad and \quad Lee: \quad E_c = 5300 f_c^{1/3} \tag{25}$$

$$Nath \quad and \quad Sarker \quad E_c = 3510\sqrt{f_c} \tag{26}$$

$$Bellum \quad et \quad al: \quad E_c = 3282\sqrt{f_c} \tag{27}$$

$$Hardjito \ et \ al: \ E_C = 2707\sqrt{f_c} + 5300 \tag{28}$$

$$Diaj - loya \ et \ al: \ E_c = 0.037 * \rho^{1.5} * \sqrt{f_c} \tag{29}$$

Where E_c is elastic modulus in MPa, ρ is the density of the concrete in kg/m^3 and f_c is characteristic strength in MPa

5. Conclusions

This conclusion is based on all results from the experimental investigation conducted on the GPC identical mix design specimens. All types of specimens cured in both types of conditions are ambient-cured and oven-cured. After the experimental investigation generates the correlation equations among the mechanical properties are as follows

- The ambient cured GPC mix specimens got higher density and drying shrinkage than the oven-cured specimens.
- The ambient cured specimens have a higher Poisson's ratio than the oven-cured specimens, whereas the oven-cured specimen's elastic modulus is higher than the ambient-cured specimens.
- The oven-cured specimens have a higher compressive strength, splitting tensile, and flexural strength than the ambient cured specimens.
- The rebound strength and UPV of oven-cured samples are more significant than the ambient-cured samples of the same GPC mix.
- The intensity of the oven-cured sample is lower than the ambient-cured sample due to the decrease of the oxide mineral components during oven-curing. However, the intensity pattern of both cured samples follows a similar trend.
- The weight preserved as a percentage of oven-cured samples is more significant than ambient-cured samples up to 800 °C. The weight preserved by ambient-cured and oven-cured samples is 92% and 94%, respectively.
- As per the experimental results, proposed the correlation equation between compressive strength and splitting tensile strength is $f_{st} = -0.92 + 0.16 f_c$, correlation equation between compressive strength and flexural strength is $f_{fs} = -1.23 + 0.19 f_c$ and correlation equation between the compressive strength and elastic modulus is $E_c = 15.08 + 0.24 f_c$

Acknowledgments

This work is supported by the Civil Engineering Department, Delhi Technological University. Moreover, Sandeep Mishra helps with XRD analysis and Ankesh Kumar and Jawed Alam for another work related to the central instrumentation facility of the Delhi Technological University.

Data availability statement

The data that support the findings of this study are available upon reasonable request from the authors.

Declarations

Funding

This project is supported by the Delhi Technological University, Delhi, India.

Conflicts of interest/competing interests

The authors declare that they have no conflicts of interest in this work.

Availability of data and material

All data analyzed related to the work were published in this article.

Code availability

Not applicable

Authors contributions

All authors were involved in every phase of the paperwork, and All authors read and approved the final manuscript.

Ethics approval

Not applicable

Consent to participate

All the authors are informed about the submission and consent to publish the manuscript.

Consent for publication

The authors give their consent for publication.

Compliance with ethical standards

All authors declared that the manuscript follows all the ethical standards.

Disclosure of potential conflicts of interest

The authors declare that they have no conflicts of interest in this work.

Research involving human participants and animals

Not applicable

Informed consent

Not applicable

ORCID iDs

Manvendra Verma https://orcid.org/0000-0002-0840-841X

References

Ahmed H Q, Jaf D K and Yaseen S A 2020 Comparison of the flexural performance and behaviour of fly-ash-based geopolymer concrete beams reinforced with CFRP and GFRP Bars *Adv. Mater. Sci. Eng.* 2020

Albitar M *et al* 2015 Assessing behaviour of fresh and hardened geopolymer concrete mixed with class-F fly ash *KSCE J. Civ. Eng.* **19** 1445–55

ASTM C 618 19 2019 *Standard Specification for Coal Fly Ash and Raw or Calcined Natural Pozzolan for Use in concrete* (ASTM International)

ASTM C138/C138M 17a 2019 Standard test method for density (unit weight) *Yield, and Air Content (Gravimetric) of Concrete* 1–6 ASTM International

ASTM C157/C157M 08 2008 Standard test method for length change of hardened hydraulic-cement mortar and concrete **08** 1–7 (Reapproved)

ASTM C403/C403M 16 2019 *Standard Test Method for Time of Setting of Concrete Mixtures by Penetration Resistance* (ASTM International) 1–7

ASTM C469/C469M 14 2019 *Standard Test Method for Static Modulus of Elasticity and Poisson's Ratio of Concrete* (ASTM International) 1–5

ASTM D934 13 2014 *Identification of Crystalline Compounds in Water-Formed Deposits by x-Ray Diffraction* (ASTM International)

Bellum R R *et al* 2019 Annales de chimie - science des matériaux empirical relationships on mechanical properties of class-f fly ash and ggbs based geopolymer concrete *Annales de Chimie - Science des Matériaux* **43** 189–97

Bondar D *et al* 2010 Engineering properties of alkali activated natural pozzolan concrete *Second International Conference On Sustainable Construction Materials And Technologies* pp 1–10 Available at: (http://www.claisse.info/Proc..htm)

Bondar D *et al* 2015 Sulfate resistance of alkali activated pozzolans *International Journal of Concrete Structures and Materials. Korea Concrete Institute* **9** 145–58

Committee A C I 2014 *Building Code Requirements for Structural Concrete (ACI 318-14)*

Concrete C 2009 *Reinforced Concrete Design in Accordance With AS* 3600–2009

Davidovits J 1989 Geopolymers and Geopolymeric Materials *J. Therm. Anal.* **35** 429–41

Davidovits J and Quentin S 1991 Geopolymers Inorganic polymeric new materials *Journal ofThamalAnalysis* **37** 1633–56

Fan F *et al* 2018 Mechanical and thermal properties of fly ash based geopolymers *Constr. Build. Mater.* (Elsevier Ltd) 160, 66–81

Finnie K S *et al* 2007 Influence of curing schedule on the integrity of geopolymers *J. Mater. Sci.* **42** 3099–106

Gardner N J and Poon S M 1976 Time and temperature effects on tensile, bond and compressive strengths *ACI Journal* **73** 405–9

Gomes R F, Dias D P and Silva F de A 2020 Determination of the fracture parameters of steel fiber-reinforced geopolymer concrete *Theor. Appl. Fract. Mech.* Elsevier **107** 102568

Hardjito D *et al* 2004 On the development of fly ash-based geopolymer concrete *ACI Mater. J.* **101** 467–72

Heah C Y *et al* 2011 Effect of curing profile on kaolin-based geopolymers *Phys. Proc.* (Elsevier B.V) 22, 305–11

International, C. of E. 1990 ceb-fip-model-code-1990-design-code.pdf.

IS 10086 1982 2008 Specification for moulds for use in tests of cement and concrete Bureau of Indian Standards, (1952)

IS 1199 1959 1959 Methods of sampling and analysis of concrete Bureau of Indian Standards, (2004)

IS 12119 1987 1999 General requirement for pan mixtures for concrete Bureau of Indian Standards, (1999)

IS 13311 (Part1) 1992 Non-destructive testing of concrete methods of test Part 1: Ultrasonic pulse velocity Bureau of Indian Standards, (2004)

IS 13311-2 1992 Method of non-destructive testing of concrete-methods of test Part 2: Rebound hammer Bureau of Indian Standards, (2004) Available at: (https://law.resource.org/pub/in/bis/S03/is.13311.2.1992.pdf)

IS 2386 (Part I) 1997 Methods of test for aggregates for concrete: I. Particle size and shape *Bureau of Indian Standards* 2386

IS 2386 (Part III) 1997 Methods of test for aggregates for concrete: II. Specific gravity, density, voids, absorption and bulking Bureau of Indian Standards, 2386(October 1963)

IS 2386 (Part IV) 1997 Methods of test for aggregates for concrete: IV. Mechanical properties Bureau of Indian Standards, 2386(March 1964)

IS 2386 (PartV) 1997 Method of test for aggregates for concrete: V. Soundness *Bureau of Indian Standards.*

IS 383 1970 1997 Specification for coarse and fine aggregates from natural sources for concrete *Bureau of Indian Standards* **1997** 1–20

IS 456 2000 2000 Plain and reinforced concrete - code of practice *Bureau of Indian Standards*, (July)

IS 516 1959 2004 Methods of test for strength of concrete Bureau of Indian Standards, (2004)

IS 5816 1999 1999 Splitting tensile strength of concrete—method of test *Bureau of Indian Standards*, (1999)

IS 7320 1974 2008 Specification for concrete slump test apparatus Bureau of Indian Standards, (2008)

IS 9103 1999 1999 Concrete admixtures-specification *Bureau of Indian Standards*, April

IS 9399 1979 1979 Specification for apparatus for flexural testing of concrete Bureau of Indian Standards, (2004)

Islam A *et al* 2015 Engineering properties and carbon footprint of ground granulated blast- furnace slag-palm oil fuel ash-based structural geopolymer concrete *Construction & Building Materials.* (Elsevier Ltd) 101, 503–21

Jang J G, Lee N K and Lee H K 2014 Fresh and hardened properties of alkali-activated fly ash/slag pastes with superplasticizers *Constr. Build. Mater.* (Elsevier Ltd) 50, 169–76

Joseph D 2015 *Geopolymer Chemistry & Applications*

Juki M I *et al* 2013 Relationship between compressive, splitting tensile and flexural strength of concrete containing granulated waste polyethylene terephthalate (PET) bottles as fine aggregate *Advanced Materials Research* **795** 356–9
Kani E N and Allahverdi A 2009 Effects of curing time and temperature on strength development of inorganic polymeric binder based on natural pozzolan *J. Mater. Sci.* **44** 3088–97
Kong D L Y and Sanjayan J G 2010 Cement and concrete research effect of elevated temperatures on geopolymer paste, mortar and concrete *Cem. Concr. Res.* (Elsevier Ltd) 40, 334–9
Krivenko P V and Kovalchuk G Y 2007 Directed synthesis of alkaline aluminosilicate minerals in a geocement matrix *J. Mater. Sci.* **42** 2944–52
Kumar R, Verma M and Dev N 2021 Iinvestigation on the effect of seawater condition , sulphate attack , acid attack , freeze—thaw condition and wetting—drying on the geopolymer concrete *Iranian Journal of Science and Technology, Transactions of Civil Engineering.* (Springer International Publishing)
Kumar R, Verma M and Dev N 2022 Investigation of fresh, mechanical and impact resistance properties of rubberized concrete *International e-Conference on Sustainable Development & Recent Trends in Civil Engineering* 88–94
Kumar, Shiv B, Mahantesh N and Amarnath K 2015 Influence of ambient curing on reinforced geopolymer concrete structural elements *International Journal of Engineering Trends and Technology* **28** 197–200
Larrard F. de 1999 *Concrete Mixture Proportioning*
Lee N K and Lee H K 2013 Setting and mechanical properties of alkali-activated fly ash/slag concrete manufactured at room temperature *Constr. Build. Mater.* (Elsevier Ltd) **47** 1201–9
Lloyd N A and Rangan B V 2010 Geopolymer concrete with fly ash *Second International Conference On Sustainable Construction Materials And Technologies, (January 2010), 1–7* Available at: (https://www.researchgate.net/publication/228825101)
Nagalia G *et al* 2016 Compressive strength and microstructural properties of fly ash—based geopolymer concrete *J. Mater. Civ. Eng.* **28** 1–11
Nath P and Sarker P K 2017 Flexural strength and elastic modulus of ambient-cured blended low-calcium fly ash geopolymer concrete *Constr. Build. Mater.* (Elsevier Ltd) 130, 22–31
Nematollahi B and Sanjayan J 2014 Effect of different superplasticizers and activator combinations on workability and strength of fly ash-based geopolymer *Mater. Des.* (Elsevier Ltd) 57, 667–72
Palacios M, Banfill P F G and Puertas F 2008 Rheology and setting of alkali-activated slag pastes and mortars: effect if organic admixture *ACI Mater. J.* **105** 140–8
Palacios M and Puertas F 2011 Effectiveness of mixing time on hardened properties of waterglass-activated slag pastes and mortars *ACI Mater. J.* **108** 73–8
Patankar S V, Ghugal Y M and Jamkar S S 2014 Effect of concentration of sodium hydroxide and degree of heat curing on fly ash-based geopolymer mortar *Indian J. Mater. Sci.* **2014** 1–6
Rovnaník P 2010 Effect of curing temperature on the development of hard structure of metakaolin-based geopolymer *Constr. Build. Mater.* (Elsevier Ltd) 24, 1176–83
Russell H G *et al* 1997 *State-of-the-Art Report on High-Strength Concrete Reported by ACI Committee* 363
Ryu G S *et al* 2013 The mechanical properties of fly ash-based geopolymer concrete with alkaline activators *Construction & Building Materials* (Elsevier Ltd) 47, 409–18
Sathonsaowaphak A, Chindaprasirt P and Pimraksa K 2009 Workability and strength of lignite bottom ash geopolymer mortar *J. Hazard. Mater.* **168** 44–50
Verma M and Dev N 2017 Review on the effect of different parameters on behavior of Geopolymer Concrete *International Journal of Innovative Research in Science, Engineering and Technology* **6** 11276–81
Verma M and Dev N 2018 Geopolymer concrete: a way of sustainable construction *International Journal of Recent Research Aspects* **5** 201–5
Verma M and Dev N 2020 Sodium hydroxide effect on the mechanical properties of flyash-slag based geopolymer concrete *Structural Concrete. fib. International Federation for Structural Concrete* **22** E368–79
Verma M and Dev N 2021a Effect of ground granulated blast furnace slag and fly ash ratio and the curing conditions on the mechanical properties of geopolymer concrete *Structural Concrete* 1–15
Verma M and Dev N 2021b Effect of liquid to binder ratio and curing temperature on the engineering properties of the geopolymer concrete *Silicon.* (Netherlands: Springer) 1–15
Verma M and Dev N 2021c Effect of SNF-based superplasticizer on physical, mechanical and thermal properties of the geopolymer concrete *Silicon.* (Netherlands: Springer) 1–11
Verma M and Nigam M 2017 Mechanical behaviour of self compacting and self curing concrete *International Journal of Innovative Research in Science, Engineering and Technology* **6** 14361–6
Xu H and Deventer J S J V 2002 Geopolymerisation of multiple minerals *Miner. Eng.* **15** 1131–9

Silicon
https://doi.org/10.1007/s12633-022-01822-4

ORIGINAL PAPER

Analog/RF Performance and Effect of Temperature on Ferroelectric Layer Improved FET device with Spacer

Yash Pathak[1] · **Bansi Dhar Malhotra**[2] · **Rishu Chaujar**[1]

Received: 1 November 2021 / Accepted: 12 March 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract

In this article, we investigated the analog performance and RF(Radio Frequency) performance of ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor (MFMIMOS) with spacer and without spacer. A proposed device MFMIMOS (spacer) with spacer enhances the ON-current (I_{on}) by 25% as compared to a without spacer device MFMIMOS, leakage current (I_{off}) reduces by almost 37%, switching ratio (I_{on}/I_{off}) enhanced by 99%, threshold voltage (V_{th}) increased by 0.29%, subthreshold swing (SS) reduced by almost 5.6% and drain induced barrier lowering (DIBL) lowed by 17.6% over MFMIMOS. We also examined the analog parameters for improving the performance of the proposed device MFMIMOS with a spacer at temperature (T) 300K such as a transconductance generation factor (TGF), transconductance (g_m), intrinsic gain (A_v), early voltage (V_{ea}), intrinsic delay (T_i), and some RF parameters gain transconductance frequency product (GTFP), cut off frequency (F_t), and gain frequency product(GFP). In the variation of temperature on MFMIMOS(spacer) we observed improved assessment at room temperature rather than other temperatures T=400K and T=500K such as threshold voltage higher by 2.4 times over temperature T=500K, switching ratio (I_{on}/I_{off}), transconductance (g_m), transconductance generation factor (TGF). All the simulated result is taken by Visual TCAD simulator with high compatibility of MFMIMOS (spacer) with spacer. Thus, the proposed device MFMIMOS (spacer) with spacer shows significantly improved analog performance than its conventional counterpart and improved performance at room temperature (T=300K) rather than other temperatures T=400K and T=500K.

Keywords MFMIMOS(spacer) · Subthreshold swing (SS) · Intrinsic Delay (T_i) · Visual TCAD · Short channel effects (SCEs)

1 Introduction

Accomplishing better speed of objective of Moore's law of transistor scaling has been the stabled at small power dissipation. The power dissipation (P) ranges with the inventory voltage(P $\tilde{V}^2$ D), with V_d generally restricted by the base subthreshold swing (SS = 60mV/dec) of an obstruction monitored Boltzmann switch [1]. As indicated by Gordon E.Moore, at almost every one of eighteen months will being twice the number of transistors in a presented unit of space [2–4] . As of late, ferroelectric (Fe-FETs) have pulled in improved consideration because of their guarantee in both memory and exchanging applications [5, 6]. Size and power consumption are the main figures of value of any advanced electronic device and the objective of past and present research endeavors is to devise forcefully scaled devices that consume ultralow power [7]. To conquer this drawback, there has been extraordinary interest in elective devices dependent on a very basic level diverse activity systems that bring about a change between the off and on condition of an electronic switch [6].

Ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor (MFMIMOS) with spacer and without spacer is

✉ Rishu Chaujar
chaujar.rishu@dtu.ac.in

Yash Pathak
yash_2k19phdap505@dtu.ac.in

Bansi Dhar Malhotra
bansi.malhotra@dce.ac.in

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

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

Published online: 14 May 2022

known as NEGATIVE capacitance FETs (NCFETs) could work at a sub-60-mV/decade subthreshold swing and a high ON-current. Presently a day [7, 8], negative capacitance marvels have been experimentally shown in various frameworks:1) isolated ferroelectric images 2) ferroelectric-dielectric bilayers 3) superlattices and ferroelectric gated transistors [9, 10]. MFMIMOS did exclude the effect of channel length scaling. Additionally, the last utilizes the Landau model just in the outside-of-plane polarization part, The Landau's parameters for the FE material are $\alpha = -1.0 \times 10^{+7}$ and $\beta = 8.9 \times 10^{+8}$ [11], which is not reasonable for huge ferroelectric thicknesses [12]. In our previous works, we examined and analyzed the execution of the MFMIMOS with lengthy channels, utilizing logical displaying [13]. MFMIMOS gives a higher ON-state flow contrasted and MFMIMOS; be that as it may, we discover this not to be the case for all the ferroelectric materials. Moreover, a definite actual understanding of the conduct of the two constructions and general performance analysis is yet missing in the writing.

In this essay, we perform a broad coordinated correlation between MFMIMOS (spacer) and MFMIMOS constructions of NCFET utilizing a conservative demonstrating approach [12]. In the MFMIMOS structure, the metal layer insert between the ferroelectric layer and the gate insulator gives an equipotential side at the interior gate [14, 15], consequently empowering the management of the ferroelectric and the fundamental MOSFET as two separate circuit elements associated by a wire [16–18].

Further, ferroelectric field effect transistors (FeFETs) stand out for new era of FET devices in view of low power loss, non-volatile read ability as a memory device, and quick operation speed [19–21]. The utilization of ferroelectric material in the gatestack of MOSFET gives better resistance to the device execution since FE material shows the conduct of negative capacitance which can be seen from their charge energy bend [22–24]. So we can utilize the ferroelectric materials in numerous potential ways to reclassify the engineering of existing MOSFET structures. For that, an appropriate illustration of FE materials is HfO_2 based FE (HfO_2FE). Essentially, we know that the HfO_2 is high-k dielectric material however when it is deposited with chemical vapor deposition, its properties show the ferroelectric conduct [11].

In this article, Ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor with spacer (MFMIMOS (spacer)) at room temperature (T=300K) enhanced the performance of the device in comparison to other MFMIMOS devices such as lower subthreshold swing, lower leakage current, lower drain induced barrier lowering, low input power consumption and control over gate channel device. A new concept of a ferroelectric layer with dielectric constant (K=33.5) has been introduced. The matrix of device performance has been extracted by using visual TCAD simulator. In this MFMIMOS (spacer), the I_{off} is lying in the request for 10^{-9} A and thus has less releasing capacity from a battery point of view, which is useful for superior performance. The MFMIMOS has upgraded the sum of Ion and great depletion channel at off-state (i.e., at V_{gs} = 0). From this examination, we become more acquainted with the device that has phenomenal execution of analog and power amplification at the point when having an analogy and RF of simulation between MFMIMOS(spacer), MFMIMOS.

In this article, We have also observed ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor with spacer (MFMIMOS (spacer)) at temperature T=300K enhanced the performance of the device in comparison to other temperatures such as T=400K and T=500K of MFMIMOS (spacer) devices such as lower subthreshold swing, lower leakage current.

2 Device Architecture

In Fig. 1(a) shows the proposed device MFMIMOS (spacer) with spacer. There are also two schematics represented by the names as Fig. 1(a) represents metal insulator metal oxide transistor with spacer (MFMIMOS (spacer)), Fig. 1 (b) shows metal insulator metal oxide transistor without spacer (MFMIMOS).

The Body's Gate length (L_b) is 60nm, Drain's / Source's length ($L_{d/s}$) is 10nm, the Spacer1's /Spacer2's length ($L_{sp1/sp2}$) is 10nm, Body's thickness (T_b) is 50nm, Drain's/ Source's thickness ($T_{d/s}$) is 3nm, Oxide's thickness (T_{ox}) is 1nm, Ferroelectric's thickness (T_{fe}) is 1nm, Gate's / Metal2's ($T_{g/m2}$) thickness is 2nm, Insulator's thickness (T_{ins}) is 2.2nm. As depicted in Table 1, the total gate length (L_g) is 20nm for MFMIMOS (spacer) and MFMIMOS. The thickness of the total gate (T_{gt}) is 10nm MFMIMOS (spacer) and MFMIMOS. Numerous classical models are utilized to clarify various ideas in the simulation work like Shockley-Read-Hall (SRH) model briefs the trap charges presence at the point of interaction, recombination impacts. The Fermi-Dirac insights are utilized for precision in the outcomes. Arora model characterized the concentration dependent mobility.

The material of nitride is used silicon nitride (Si_3N_4). These compounds have high melting points, are extremely hard, and are usually opaque materials that have metallic lustre and high conductivities. Silicon nitride offers a number of benefits, such as high temperature capability, low friction, high wear and chemical resistance, and high electrical resistivity. It is nonmagnetic, has low specific

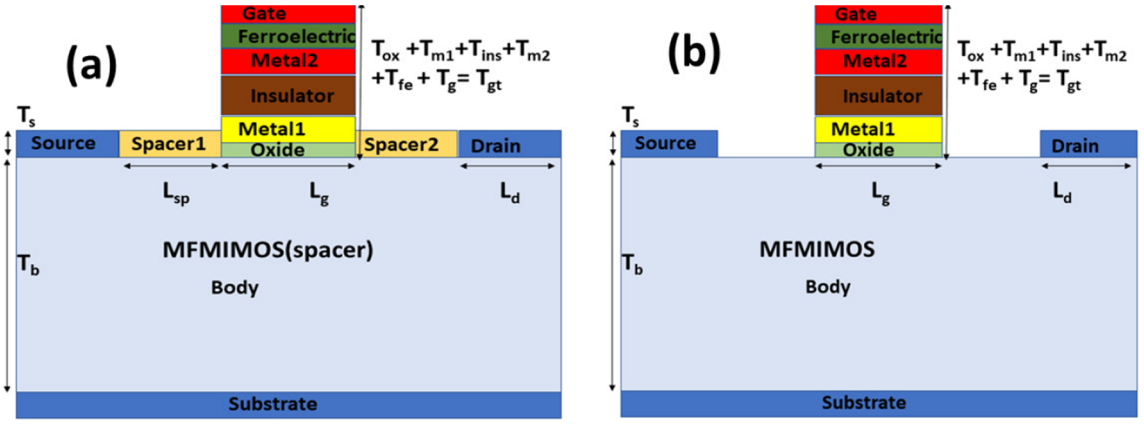


Fig. 1 The two different MFMIMOS schematics structure with using new concept of ferroelectric layer (a) MFMIMOS (spacer) with spacer (b) MFMIMOS

weight, and offers long service life. Detection of nitride is vital for human and environmental protection [25]. As shown in Fig. 1, the material of the Body is Si (silicon), the material of Drain / Source /Substrate/Gate is Al (aluminum), the material of Spacer1 /Spacer2 is Nitride, the material of Oxide is $Si\,O_2$ (silicon dioxide), Material of Ferroelectric is $Hf\,O_2$FE (a new idea of a ferroelectric layer with hafnium dioxide), and material used for Metal1, Metal2 and Insulator is Au, NPoly Si and $Hf\,O_2$ respectively. A proposed device MFMIMOS (spacer) involve of gate, ferroelectric, metal, insulator, metal, silicon dioxide, silicon from top to bottom. The doping concentration of Body (N_b) and silicon substrate is 1×10^{16} cm^{-3} with Uniform profile and acceptor type, the doping concentrating of Drain /Source ($N_{d/s}$) is 1×10^{20} cm^{-3} with Gaussian profile and type of donor and the doping concentration of threshold voltage (N_{vt}) is 1×10^{18} cm^{-3} with Gaussian profile and acceptor type. The work function is retained at $4.5eV$ for a gate of metal.

3 Method of Simulation

All the simulations are taken by Visual TCAD simulator. Comparison of two devices i.e MFMIMOS (spacer), and MFMIMOS have been carried out. Drain voltage, V_d (V) is fixed in the entire process at 0.3V. Gate voltage, V_{gs} is different from 0V to 3V. temperature (T) is fixed during the the first case device simulation at 300K (room temperature). In the second case variation of temperature on MFMIMOS (spacer). We have disregarded the connection among ferroelectric dipoles, which is a sensible suspicion for g_d: HfO_2 material [26, 27]. Distinctive limit conditions are applied to carry out the MFMIMOS and MFMIMOS designs. In the MFMIMOS structure, congruity of typical parts of displacement vectors is utilized at the oxide/spacer ferroelectric limits. For the MFMIMOS structure, the inside metal gate goes about as a coasting terminal. We utilize visual TCAD Multiphysics to play out every single mathematical reproduction.

Table 1 These parameters are used for the simulation of MFMIMOS (spacer)

Parameter	MFMIMOS (spacer)	MFMIMOS
Gate length (L_g)	20 nm	20 nm
Oxide Thickness (T_{ox})	1 nm	1 nm
Drain/Source length ($L_{d/s}$)	10 nm	10 nm
Drain/ Source thickness ($T_{d/s}$)	3 nm	3 nm
Body thickness (T_b)	50 nm	50 nm
Ferroelectric thickness (T_{fe})	1 nm	1 nm
Insulator thickness (T_{ins})	2.2 nm	2.2 nm
Concentration of Body (N_b)	1×10^{16} cm^{-3}	1×10^{16} cm^{-3}
Concentration of Drain/ Source ($N_{d/s}$)	1×10^{20} cm^{-3}	1×10^{20} cm^{-3}
Concentration of threshold voltage (N_{vt})	1×10^{18} cm^{-3}	1×10^{18} cm^{-3}
Length of Spacer1/Spacer2 ($L_{sp1/sp2}$)	10nm	10nm

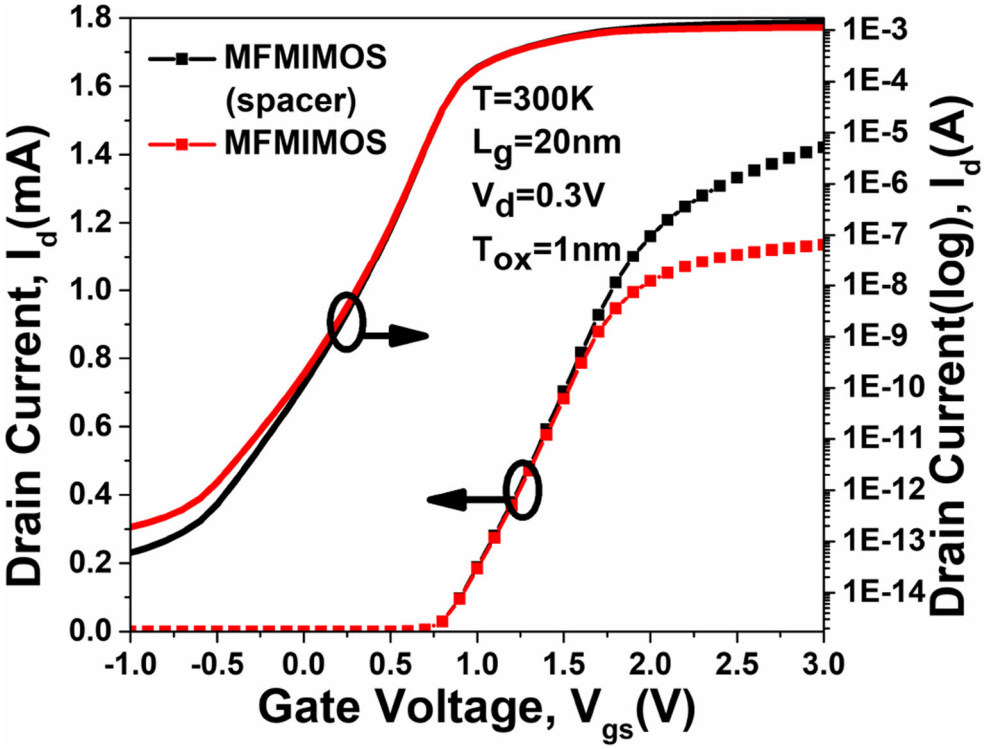


Fig. 2 The curve of drain current vs gate voltage on the left side of the graph and right side of the y-axis is log scale drain current for MFMIMOS (spacer), MFMIMOS at $V_d = 0.3V$

4 Result & Discussion

4.1 Device Scalability

Figure 2 reveals the curve between drain current, I_d (A) in linear scale and drain current [28, 29], I_d (A) in log scale vs gate voltage, V_{gs} (V) for ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor with spacer (MFMIMOS (spacer)) at temperature T=300K, and MFMIMOS at V_d=0.3V, channel length, L_g=20nm, T=300K, and oxide thickness T_{ox}=1nm. It reveals thet Off state current lower for MFMIMOS (spacer) in the log drain current in negative bias also. The drain current of MFMIMOS (spacer) is better than the MFMIMOS configurations with increase in gate voltage V_{gs} (V) that reflects the Superior gate coupling capacitance and reduced leakage current I_{off} in the left side of the graph.

Figure 3 shows that the threshold voltage, V_{th} (V) of MFMIMOS (spacer) is enhanced than other MFMIMOS device configurations at drain voltage, V_d =0.3V due to increased gate control over the channel and structure MFMIMOS (spacer) is the better shield of drain-side potential so that improved the character of the threshold voltage. The switching ratio, I_{on}/I_{off} of MFMIMOS (spacer) is higher than the other MFMIMOS structures at $L_g = 20nm$, $T_{ox} = 1nm$, $T = 300K$, and $V_d = 0.3V$, decreased leakage current, I_{off} (A), and high switching speed for MFMIMOS (spacer) structure is also obtained. The log current for measuring of V_{th} (I_{th}) is fixed $1 \times 10^{-7}A$. The formula is used for calculating V_{th} here, V_1, I_2, I_1, V_2 is $\log(v_2)$, $\log(i_2 \times W/L)$, $\log(i_1 \times W/L)$, $\log(v_2)$. The expression $V_g = v_2 - -(I_2 - \log(I_{th})) \times (v_2 - v_1)/(I_2 - I_1)$.

4.2 Analog Analysis

In this subsection, analog performance metrics are discussed in terms of transconductance (g_m), transconductance generation factor (TGF), output conductance [30], early voltage (V_{ea}), intrinsic gain (A_v). Transconductance, transconductance generation factors are higher for MFMIMOS (spacer).

Figure 4 reveals that transconductance is a parameter of the shift in drain current (I_d) to shift in gate voltage (V_{gs}) at drain voltage (0.3V) at T=300K. The transconductance curve of MFMIMOS (spacer) is higher than the MFMIMOS structures. Thus, g_m has taken from the curve drain current vs gate voltage that reflects enhanced gate control and lowered short channel effects (SCEs) and also increases average carrier velocity, upgraded electron mobility with increase transconductance [31].

$$g_{\mathrm{m}} = \frac{\partial I_{\mathrm{d}}}{\partial V_{\mathrm{gs}}} \tag{1}$$

$$TGF = \frac{\partial g_{\mathrm{m}}}{\partial I_{\mathrm{d}}} \tag{2}$$

$$g_{\mathrm{d}} = \frac{\partial I_{\mathrm{d}}}{\partial V_{\mathrm{d}}} \tag{3}$$

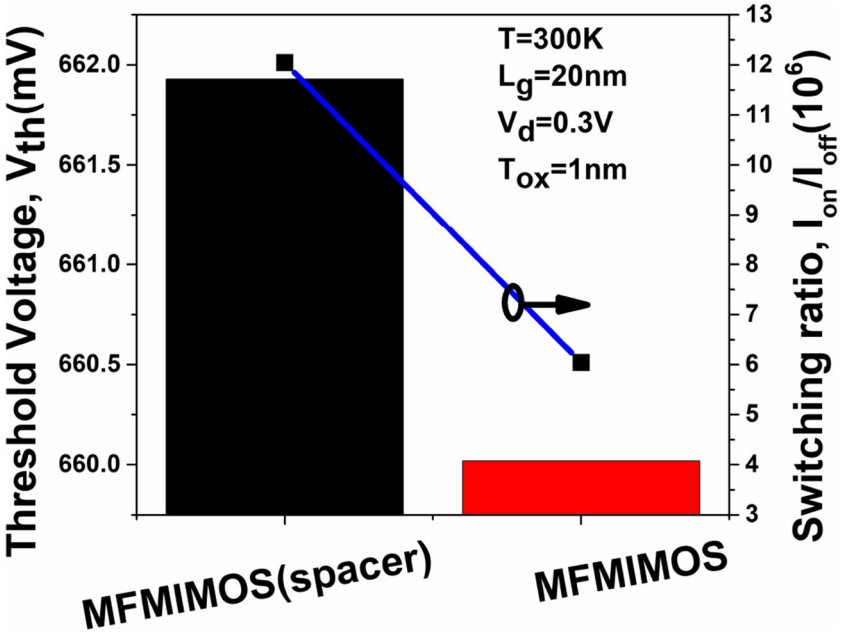


Fig. 3 The bar graph indicates for two configurations at V_d=0.3V (a) threshold voltage, V_{th} (V) (b) switching ratio

$$V_{\mathrm{ea}} = \frac{\partial I_{\mathrm{d}}}{\partial g_{\mathrm{d}}} \tag{4}$$

$$A_{\mathrm{v}} = \frac{\partial g_{\mathrm{m}}}{\partial g_{\mathrm{d}}} \tag{5}$$

Transconductance generation factor (TGF) is gain generated per unit power loss. TGF of MFMIMOS (spacer) is increased than MFMIMOS configurations, close to ideally $50V^{-1}$ and subthreshold swing ideally close to 60mV/decade. The device operation at low supply voltage performs more efficiency for a high value of TGF. (b) reflects the output characteristics for different drain current at constant gate voltage= 0.7V it gives us better than other gate voltage such as V_{gs}=0.75V and 0.8V. The changes of TGF occur in the subthreshold region. The region for high TGF and g_m due to high drain current. Figure 5 represents the graph of drain current, I_d (A) and output conductance, g_d (nS) vs drain voltage, V_d (V) at gate voltage V_{gs} (0.8) oxide thickness 1nm and channel length $20nm$. Output conductance is the driving capacity of the device as characterized in (3) [32] mirrors the locale of device activity.

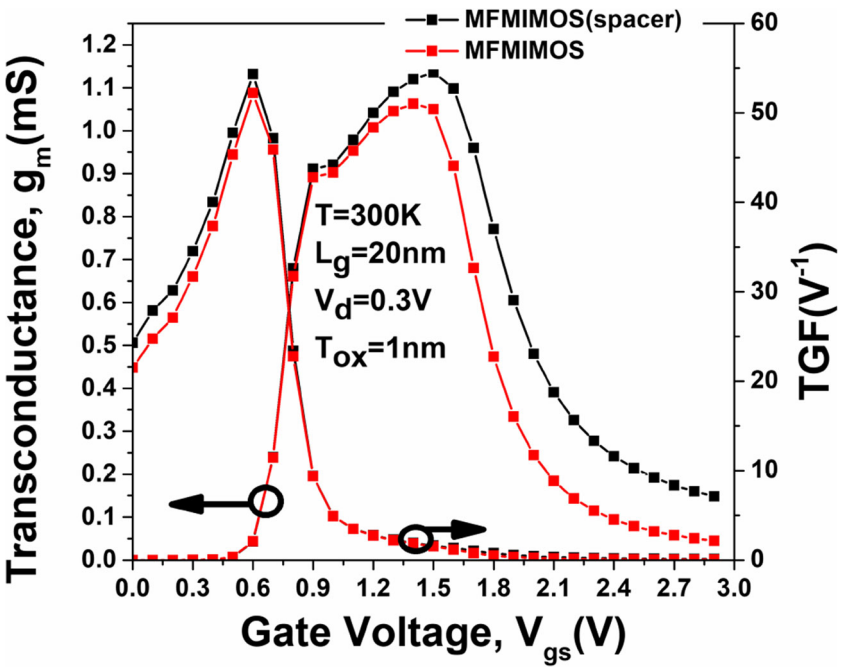


Fig. 4 The curve of different structures at V_d=0.3V (a) transconductance and gate voltage (b) transconductance generation factor vs gate voltage

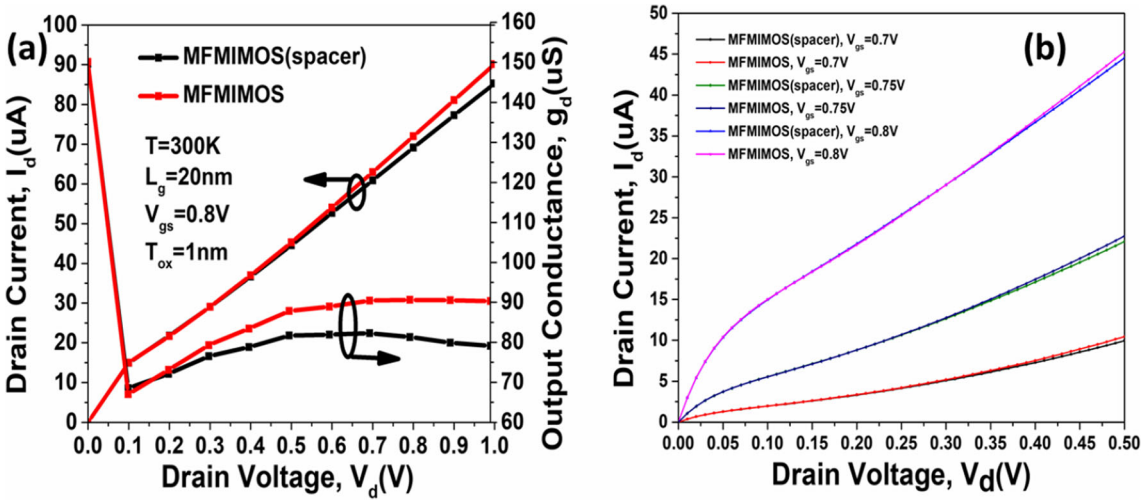


Fig. 5 (a)The curve of drain current and output conductance vs drain current for different structures at V_{gs}=0.8V.(b) Plot for various constant gate voltage

At first, a high output conductance in the linear region is observed with increasing drain voltage beyond pinch-off voltage owing to drain induced barrier lowering (DIBL) as well as channel length modulation (CLM) [33]. The g_d twist of MFMIMOS (spacer) is lower than MFMIMOS that shows upgraded gate controllability and stifle short direct impacts as portrayed in Fig. 5 that reflects improved output obstruction.

Figure 6 shows that the bar graph of drain induced barrier lowering (DIBL) for a different configuration at T=300K. It is a short channel effect (SCEs), the length of the channel decreasing means electrons easily move from source to drain. DIBL of MFMIMOS (spacer) is lower than MFMIMOS structure that reflects reduced leakage current, improved SCEs. Subthreshold swing (SS) is also a vital parameter of SCEs, ideally SS 60mV/decade. SS of MFMIMOS (spacer) is lowering than MFMIMOS that reflect enhanced gate coupling capacitance with high permittivity (K), increase gate control over the channel.

Figure 7(a) reveals that the curve of early voltage and intrinsic gain vs gate voltage for different configurations at V_d=0.8, oxide thickness 1nm, channel length 20nm [34–37]. These formulae are evaluated by using (4) and (5). The proposed device MFMIMOS (spacer) metrics are higher

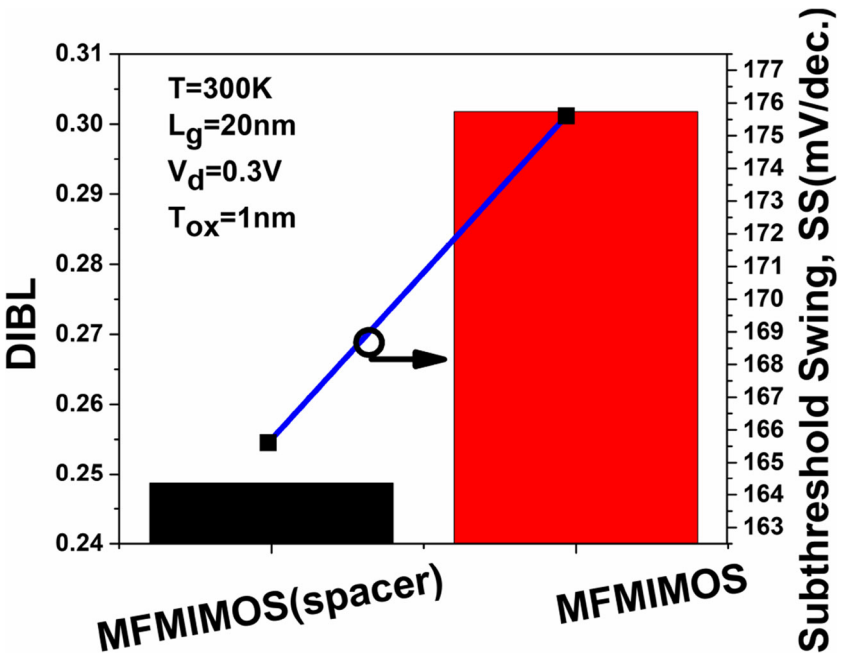


Fig. 6 Graph of different structures at V_d=0.3V (a) drain induced barrier lowering, DIBL in the left side (b) Subthreshold swing in right side

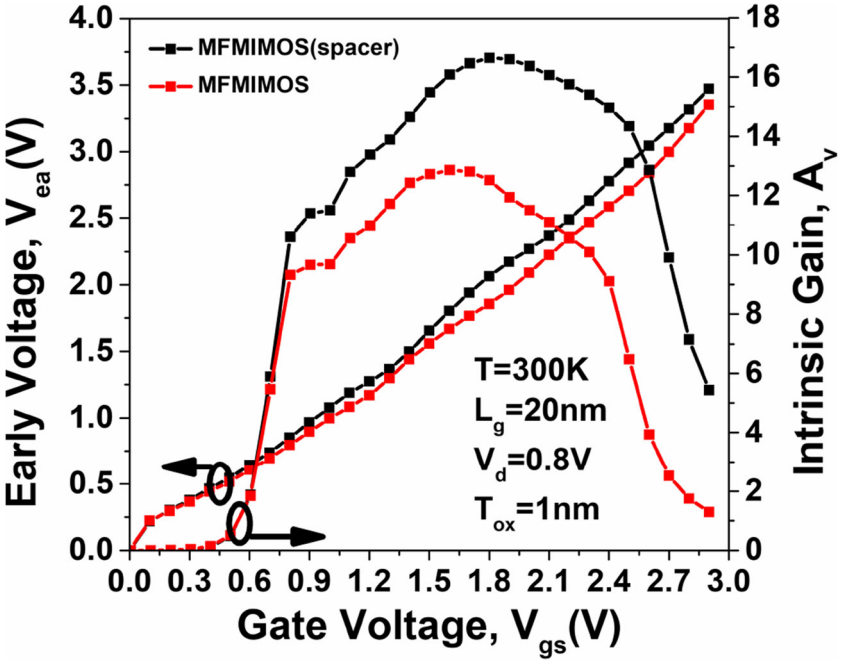


Fig. 7 Graph for different structures at V_d=0.8v (a) Variation of early voltage and intrinsic gain vs gate voltage at T=300K

than other conventional devices that indicate short channel effects (SCEs). The higher value of intrinsic gain is obtained from the high value of transconductance, and lower output conductance as depicted in Table 2.

4.3 RF Analysis

In the RF (Radio Frequency) application, the parameter of RF is gate capacitance (C_{gg}), gate charge (Q_g), unity gain cutoff frequency (F_t), gain frequency product (GFP), and gain transconductance frequency product (GTFP). etc. Fig. 8(a). represents the variation of gate capacitance, C_{gg} (F/um) vs gate voltage for all configurations at drain voltage (0.3V), channel length (20nm), and oxide thickness (1nm). The gate capacitance curve of MFMIMOS (spacer) is higher than other structures that reflect an enhanced number of stored charges.

Figure 8 indicates that gate capacitance and cutoff frequency [21, 29] of MFMIMOS (spacer) is lower than MFMIMOS configuration that reveals enhanced gate capacitance and increased transconductance as depicted in (6). Intrinsic delay of MFMIMOS (spacer) is lower than other structures by using the (7) on the right side of the graph [32]. It is the time interval carried by the device to gives output after applying input that reflects less input power consumption [32–39]. The gate charge of MFMIMOS (spacer) is higher than other structures that reflects high gate capacitance.

$$F_{\mathrm{t}} = \frac{g_{\mathrm{m}}}{2\pi(C_{\mathrm{gg}})} \tag{6}$$

$$T_{\mathrm{i}} = \frac{C_{\mathrm{gg}} \times V_{\mathrm{d}}}{(I_{\mathrm{on}})} \tag{7}$$

$$GFP = \left(\frac{g_{\mathrm{m}}}{g_{\mathrm{d}}}\right) \times F_{\mathrm{t}} \tag{8}$$

$$GTFP = \left(\frac{g_m}{g_d}\right) \times \left(\frac{g_m}{I_d}\right) \times F_t \tag{9}$$

Figure 9 shows that the electric field I contour Electric field contour with MFMIMOS (spacer) from 0 to 8.661E+06 and MFMIMOS from 0 to 8.404E+06 at $V_d = 0.3V$, T=300K, channel length 20nm, oxide thickness 1nm.

Table 2 Parameter and value of leakage current, SS, DIBL for different configuration, T_{ox}=1nm, T=300K and L_g=20nm

Parameter	MFMIMOS (spacer)	MFMIMOS
I_{on} (A)	0.001421	0.001135
I_{off} (A)(10^{10})	1.18	1.88
DIBL	0.248673	0.30171177
SS (mV/dec)	165.607	175.606
I_{on}/I_{off}(10^6)	12.0424	6.0372
$V_{th}(mV)$	661.929	660.0196

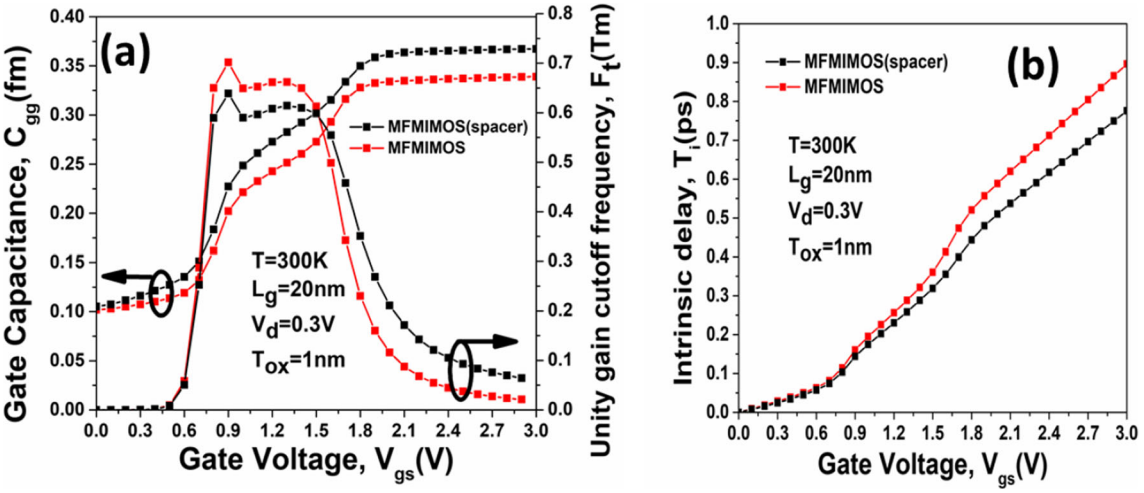


Fig. 8 Variation for different structures with V_d=0.3V and L_g = 1GHz (a) Curve of gate capacitance vs gate voltage and Curve of cutoff frequency vs gate voltage (b) Curve of intrinsic delay vs gate voltage at T=300K

The electric field is high in the area of contour from gate to drain. Area of electric field contour is higher than other structure that reflects reduced leakage current I_{off}.

Figure 10 reflects the graph for different structures at V_d=0.8 (a) Variation of gain frequency product and gain transconductance frequency gain product vs gate voltage at T=300K. Gain frequency Product is a fundamental boundary utilized in high-recurrence applications, as determined in (8) [40, 41]. It is seen from Fig. 10, GFP increments as the gate voltage higher and afterward accomplishes a greatest top prior to tumbling to a consistent worth in the saturation locale. MFMIMOS (spacer) design records the most elevated worth of GFP because of the further developed worth of output transconductance and transconductance with an insignificant abatement in cut-off frequency. GTFP most noteworthy worth is acquired for MFMIMOS (spacer) structure because of the upgraded worth of g_m, TGF, and decreased worth of g_d with practically comparable F_t. In this manner, MFMIMOS (spacer) is the best appropriate design as far as speed, transconductance, and gain.

4.4 Effect of temperature on MFMIMOS (spacer)

In this section, we observed the variation of temperature on MFMIMOS (spacer) such as T=300K, 400K and 500K drain voltage 0.3V, gate length 20nm and oxide thickness 1nm. In Fig. 11 exhibits graph for different temperature such that T=300K, 400K and 500K at V_d=0.3 (a) Drain current and log drain current vs gate voltage for MFMIMOS (spacer). The drain current of MFMIMOS (spacer) at temperature T=300K is higher than the other temperature with increase in gate voltage V_{gs} (V) that reflects the improved gate coupling capacitance and less leakage current

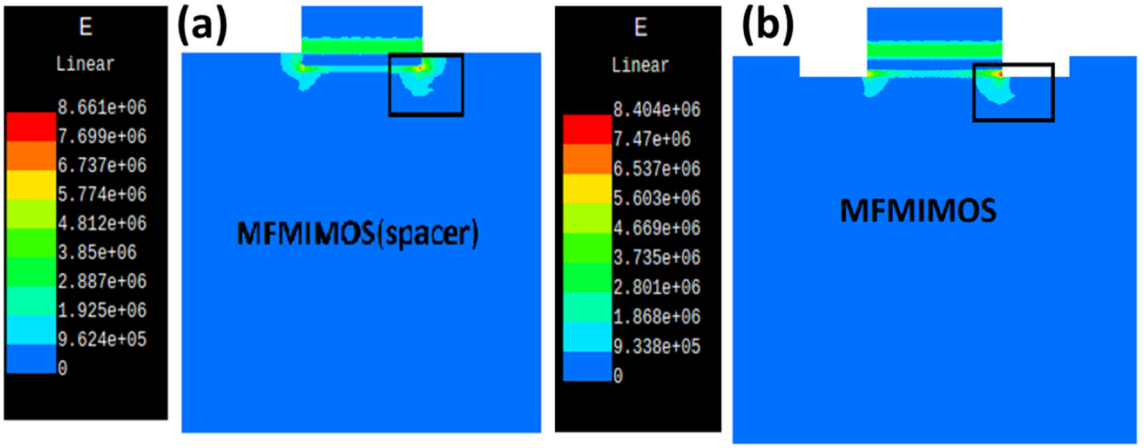


Fig. 9 Electric field contour with MFMIMOS (spacer) from 0 to 8.661E+06 and MFMIMOS from 0 to 8.404E+06 at V_d=0.3V and T=300K. (a) MFMIMOS (spacer) (b) MFMIMOS

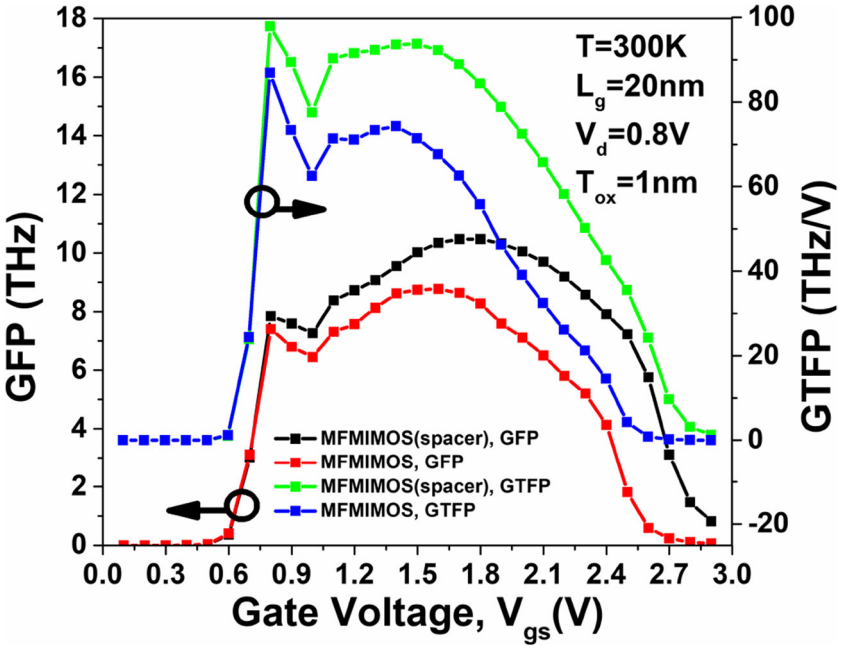


Fig. 10 Graph for different structures at V_d=0.8v (a) Variation of gain frequency product and gain transconductance frequency gain product vs gate voltage at T=300K

I_{off} in the left side of the graph. Figure 12 shows that the threshold voltage, V_{th} (V) of MFMIMOS (spacer) at temperature T=300K is enhanced than other temperature T=400K and T=500K configurations at drain voltage, V_d =0.3V due to increased gate control over the channel and structure MFMIMOS (spacer) at temperature T=300K is the better shield of drain-side potential so that improved the character of the threshold voltage. The switching ratio, I_{on}/I_{off} of MFMIMOS (spacer) temperature T=300K is higher than the other temperature T=400K and T=500K structures at L_g=20nm, T_{ox}=1nm, and V_d=0.3V, decreased leakage current, I_{off} (A), and high switching speed for MFMIMOS (spacer) structure is also obtained as depicted in Table 3.

Figure 13 reveals that transconductance is a parameter of the shift in drain current (I_d) to shift in gate voltage (V_{gs}) at drain voltage (0.3V). The transconductance curve of MFMIMOS (spacer) at temperature T=300K is higher than other temperature T=400K and T=500K structures. Thus, g_m has taken from the curve drain current vs gate voltage that reflects enhanced gate control and lowered short channel effects (SCEs) and also increases average carrier velocity, upgraded electron mobility with increase transconductance [31]. Transconductance generation factor

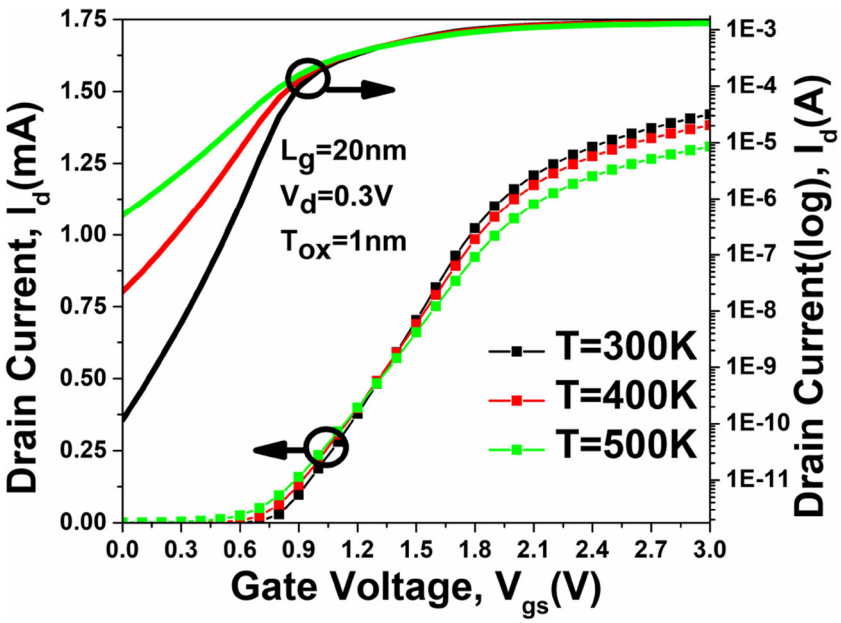


Fig. 11 Graph for different temperature such that T=300K, 400K and 500K at V_d=0.3 (a) Drain current and log drain current vs gate voltage for MFMIMOS (spacer)

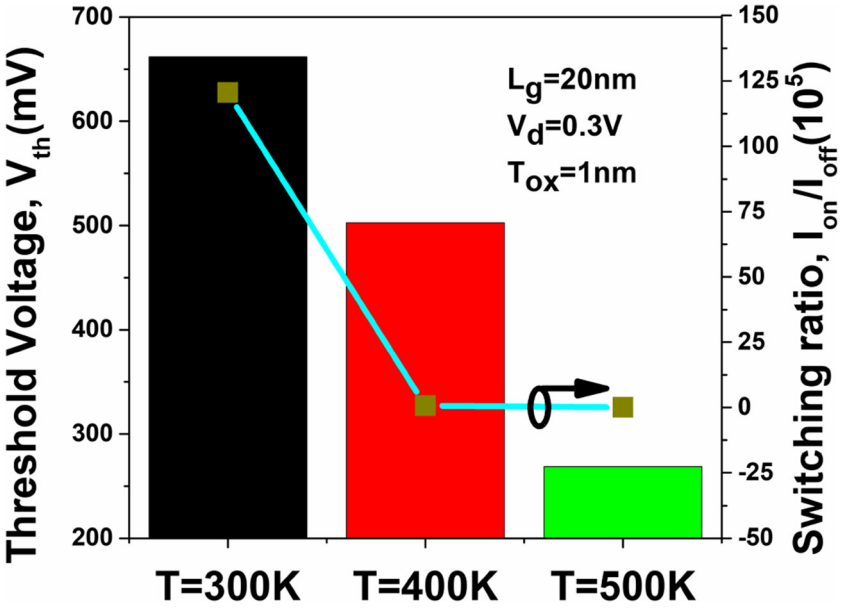


Fig. 12 Graph for different temperature such that T=300K, 400K and 500K at V_d=0.3 (a) threshold voltage and switching ratio vs gate voltage for MFMIMOS (spacer)

Table 3 Parameter and value of leakage current for different temperature at MFMIMOS (spacer) T_{ox}=1nm, and L_g=20nm

Parameter	T=300K	T=400K	T=500K
I_{on} (A)	0.001421	0.001383	0.001309
I_{off} (A)(10^{10})	1.18	220	5040
I_{on}/I_{off}(10^6)	12.0424	00.0629	0.002597
$V_{th}(mV)$	661.929	502.534	268.78

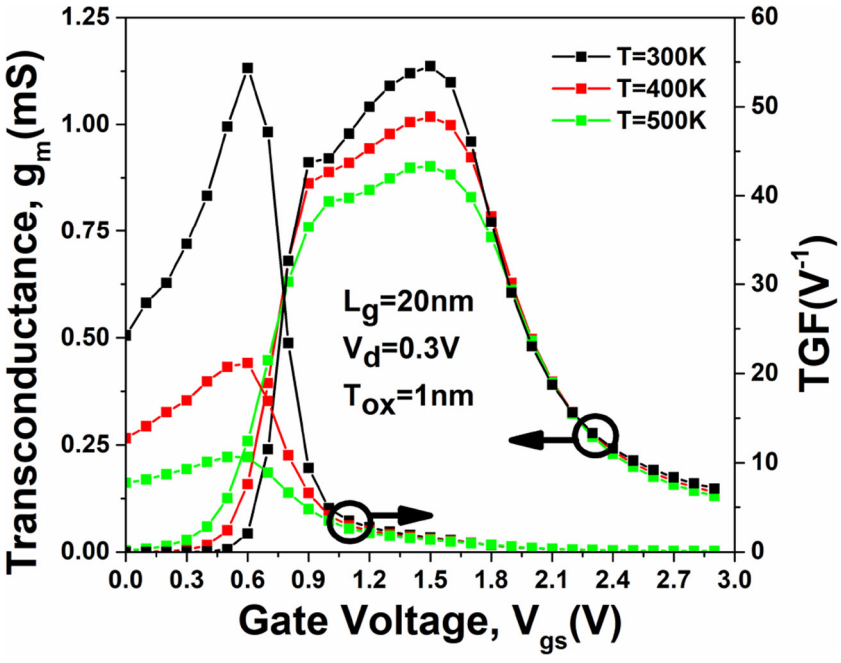


Fig. 13 Graph for different temperature such that T=300K, 400K and 500K at V_d=0.3 (a) transconductance and transconductance generation factor vs gate voltage for MFMIMOS (spacer)

(TGF) is gain generated per unit power loss. TGF of MFMIMOS (spacer) at temperature T=300K is increased than temperature T=400K and T=500K configurations, close to ideally $50V^{-1}$ and subthreshold swing ideally close to 60mV/decade. The device operation at less supply voltage performs high efficiency for a high value of TGF. The changes of TGF occur in the subthreshold region. The region for high TGF and g_m due to high drain current.

5 Conclusion

In this work, we proposed ferroelectric layer improved Field Effect Transistor device that is metal ferroelectric metal insulator metal oxide transistor (MFMIMOS) with spacer and without spacer. All the simulations are brought by a visual TCAD simulator. MFMIMOS (spacer) indicates significantly improved analog results than MFMIMOS configurations such as switching ratio increased by around 2 times of MFMIMOS, subthreshold swing reduced by 1.06 times of MFMIMOS, DIBL decreased by 1.21 times of MFMIMOS that parameters reflect a short channel effect and reduced leakage current. The structure of MFMIMOS (spacer) enhanced the properties of other analog application such as g_m, TGF, A_v, and V_{ea} increases, respectively. Furthermore. RF application is also improved such as F_t reduces, C_{gg}, GFP and GTFP. In the case of effect on temperature of MFMIMOS (spacer) at T=300K reveals more improved result than other temperature such that T=400K and T=500K. At temperature T=300K enhanced performance such as leakage current, switching ratio, transconductance TGF. Thus, this article indicates MFMIMOS (spacer) at temperature T=300K can be considered as an improved variant of analog and RF parameters in analogy to its conventional counterpart. The device has a wide variety of applications to high speed ICs.

Acknowledgements The authors would like to thank Microelectronics Lab and Nano Bioelectronics Lab, Delhi Technological University for giving essential facilities.

Author Contributions All authors added to the study's understanding and design.

Funding Yash is grateful to the Council of Scientific Industrial Research (C.S.I.R.) for an award of CSIR-JRF (08/133(0050)/2020-EMR-I). BDM thanks the Science & Engineering Research Board (DST-SERB), Govt. of India for the award of Distinguished Fellowship (SB/DF/011/2019)

Availability of Data & Material The creators expressed above have every one of the suitable data associated with this examination work and will be dedicated to uncovering that they will be addressed to do as such future.

Declarations

The authors have found all the Moral Standards and will intended to follow them in the future.

Consent for Publication Since the related research document is created on the 'no-life science journal.' Therefore, 'Non Applicable' at this point. Still, the authors have turned over all journal guidelines and consent to the agencies for additional processing.

Conflict of Interests The writers broadcast that they have no known dispute of individual connections or interests that might have arisen to impact the work depicted in this article.

References

1. Karda K., Jain A., Mouli C., Alam M. A. (2015) An anti-ferroelectric gated landau transistor to achieve sub-60 mv/dec switching at low voltage and high speed. Applied Physics Letters 106(16):163501
2. Kumar B., Chaujar R. (2021) Analog and rf performance evaluation of junctionless accumulation mode (jam) gate stack gate all around (gs-gaa) finfet. Silicon 13(3):919–927
3. Waldrop M. M. (2016) More than moore. Nature 530(7589):144–148
4. Pathak Y., Malhotra B. D., Chaujar R. (2021) Tcad analysis and simulation of double metal negative capacitance fet (dm ncfet). In: 2021 Devices for Integrated Circuit (DevIC), pp. 224–228, IEEE
5. Lee M., Chen P.-G., Liu C., Chu K., Cheng C.-C., Xie M.-J., Liu S.-N., Lee J.-W., Huang S.-J., Liao M.-H., et al. (2015) Prospects for ferroelectric hfzrox fets with experimentally cet= 0.98 nm, ssfor= 42mvdec, ssrev= 28mvdec, switch-off-lt; 0.2 v, and hysteresis-free strategies. In: 2015 IEEE International Electron Devices Meeting (IEDM), pp. 22–5 IEEE
6. Moselund K., Bouvet D., Pott V., Meinen C., Kayal M., Ionescu A. (2008) Punch-through impact ionization mosfet (pimos): From device principle to applications. Solid-state electronics 52(9):1336–1344
7. Lee M., Fan S.-T., Tang C.-H., Chen P.-G., Chou Y.-C., Chen H.-H., Kuo J.-Y., Xie M.-J., Liu S.-N., Liao M.-H., et al. (2016) Physical thickness 1. x nm ferroelectric hfzrox negative capacitance fets. In: 2016 IEEE International Electron Devices Meeting (IEDM), pp. 12–1, IEEE
8. Ahmed H. M., Chaujar R. et al (2020) Rapid detection of biomolecules in a dielectric modulated gan moshemt. Journal of Materials Science: Materials in Electronics 31(19):16609–16615
9. Lin C.-I., Khan A. I., Salahuddin S., Hu C. (2016) Effects of the variation of ferroelectric properties on negative capacitance fet characteristics. IEEE transactions on electron devices 63(5):2197–2199
10. Bidenko P., Lee S., Han J.-H., Song J. D., Kim S.-H. (2018) Simulation study on the design of sub-kt/q non-hysteretic negative capacitance fet using capacitance matching. IEEE Journal of the Electron Devices Society 6:910–921
11. Awadhiya B., Kondekar P. N., Yadav S., Upadhyay P. (2021) Insight into threshold voltage and drain induced barrier lowering in negative capacitance field effect transistor. Transactions on Electrical and Electronic Materials 22(3):267–273
12. Pahwa G., Dutta T., Agarwal A., Chauhan Y. S. (2018) Physical insights on negative capacitance transistors in nonhysteresis and hysteresis regimes: Mfmis versus mfis structures. IEEE Transactions on Electron Devices 65(3):867–873
13. Pahwa G., Agarwal A., Chauhan Y. S. (2018) Numerical investigation of short-channel effects in negative capacitance mfis and mfmis transistors: Subthreshold behavior. IEEE Transactions on Electron Devices 65(11):5130–5136
14. Vaithiyanathan D., Raj M. B., Pushpa S. E. P., Seetharaman R. (2017) Performance analysis of finfet and negative capacitance fet

over 6t sram. In: 2017 IEEE International Conference on Circuits and Systems (ICCS), pp. 196–201, IEEE
15. Goswami L. P., Maity S., Mandal D., Vashistha A., Das A. (2021). Ponderomotive force driven mechanism for electrostatic wave excitation and energy absorption of electromagnetic waves in overdense magnetized plasma Plasma Physics Controlled Fusion
16. Pahwa G., Dutta T., Agarwal A., Chauhan Y. S. (2017) Compact model for ferroelectric negative capacitance transistor with mfis structure. IEEE Transactions on Electron Devices 64(3):1366–1374
17. Luong A.-D., Roy I., Malhotra B. D., Luong J. H. (2021). Analytical and biosensing platforms for insulin: A review. Sensors and Actuators Reports 100028
18. Nara S., Kandpal R., Jaiswal V., Augustine S., Wahie S., Sharma J. G., Takeuchi R., Takenaka S., Malhotra B. D. (2020) Exploring providencia rettgeri for application to eco-friendly paper based microbial fuel cell. Biosensors and Bioelectronics 165:112323
19. Kim J. Y., Choi M.-J., Jang H. W. (2021) Ferroelectric field effect transistors: Progress and perspective. APL Materials 9(2):021102
20. Lederer M., Kämpfe T., Ali T., Müller F., Olivo R., Hoffmann R., Laleni N., Seidel K (2021) Ferroelectric field effect transistors as a synapse for neuromorphic application. IEEE Transactions on Electron Devices 68(5):2295–2300
21. Pathak Y., Malhotra B. D., Chaujar R. (2021) A numerical study of analog parameter of negative capacitance field effect transistor with spacer. In: 2021 7th International Conference on Signal Processing and Communication (ICSC), pp. 277–281, IEEE
22. Mann R., Chaujar R. (2021) Tcad investigation of ferroelectric based substrate mosfet for digital application
23. Kim H. J., Park M. H., Kim Y. J., Lee Y. H., Moon T., Do Kim K., Hyun S. D., Hwang C. S. (2016) A study on the wake-up effect of ferroelectric hf 0.5 zr 0.5 o 2 films by pulse-switching measurement. Nanoscale 8(3):1383–1389
24. Sharma S., Chaujar R. (2021) Band gap and gate metal engineering of novel hetero-material inas/gaas-based jltfet for improved wireless applications. Journal of Materials Science: Materials in Electronics 32(3):3155–3166
25. Zhou Y., Ma M., He H., Cai Z., Gao N., He C., Chang G., Wang X., He Y. (2019) Highly sensitive nitrite sensor based on aunps/rgo nanocomposites modified graphene electrochemical transistors. Biosensors and Bioelectronics 146:111751
26. Hoffmann M., Pešić M., Chatterjee K., Khan A. I., Salahuddin S., Slesazeck S., Schroeder U., Mikolajick T. (2016) Direct observation of negative capacitance in polycrystalline ferroelectric hfo2, vol 26
27. Sarinont T., Katayama R., Wada Y., Koga K., Shiratani M. (2017) Plant growth enhancement of seeds immersed in plasma activated water. Mrs Advances 2(18):995–1000
28. Gupta N., Chaujar R. (2016) Optimization of high-k and gate metal workfunction for improved analog and intermodulation performance of gate stack (gs)-gewe-sinw mosfet, vol 97
29. Song L., Fu K., Zhao J., Yu G., Hao R., Fan Y., Cai Y., Zhang B. (2018) Degradation of algan/gan metal–insulator–semiconductor high electron mobility transistors under off-state electrical stress. Journal of Vacuum Science & Technology B, Nanotechnology and Microelectronics: materials, Processing, Measurement, and Phenomena 36(4):042201
30. Mohapatra S., Pradhan K., Artola L., Sahu P. (2015) Estimation of analog/rf figures-of-merit using device design engineering in gate stack double gate mosfet. Materials Science in Semiconductor Processing 31:455–462
31. Narendar V., Girdhardas K. A. (2018) Surface potential modeling of graded-channel gate-stack (gcgs) high-k dielectric dual-material double-gate (dmdg) mosfet and analog/rf performance study. silicon 10(6):2865–2875
32. Sharma M., Chaujar R. (2020) Impact of graded back-barrier on linearity of recessed gate inaln/gan hemt. In: 2020 IEEE VLSI DEVICE CIRCUIT AND SYSTEM (VLSI DCS), pp. 154–158, IEEE
33. Amin S. I., Sarin R. (2015) Charge-plasma based dual-material and gate-stacked architecture of junctionless transistor for enhanced analog performance. Superlattices and Microstructures 88:582–590
34. Madan J., Gupta R., Chaujar R. (2017) Performance investigation of heterogeneous gate dielectric-gate metal engineered–gate all around-tunnel fet for rf applications. Microsystem Technologies 23(9):4081–4090
35. Chaujar R., Kaur R., Saxena M., Gupta M., Gupta R. (2008) Laterally amalgamated dual material gate concave (l-dumgac) mosfet for ulsi. Microelectronic engineering 85(3):566–576
36. Kumar A., Tripathi M., Chaujar R. (2018) Comprehensive analysis of sub-20 nm black phosphorus based junctionless-recessed channel mosfet for analog/rf applications. Superlattices and Microstructures 116:171–180
37. Madan J., Gupta R., Chaujar R. (2017) Mathematical modeling insight of hetero gate dielectric-dual material gate-gaa-tunnel fet for vlsi/analog applications. Microsystem Technologies 23(9):4091–4098
38. Sharma D., Vishvakarma S. K. (2015) Analyses of dc and analog/rf performances for short channel quadruple-gate gate-all-around mosfet. microelectronics journal 46(8):731–739
39. Kashyap M. P., Chaujar R. (2021) Gate oxide variability analysis of a novel 3 nm truncated fin–finfet for high circuitry performance. Silicon 13(9):3249–3256
40. Mohapatra S., Pradhan K., Artola L., Sahu P. (2015) Estimation of analog/rf figures-of-merit using device design engineering in gate stack double gate mosfet, vol 31
41. Mulaosmanovic H., Breyer E. T., Dünkel S., Beyer S., Mikolajick T., Slesazeck S. (2021) Ferroelectric field-effect transistors based on hfo2: a review Nanotechnology

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

Eur. Phys. J. Plus (2022) 137:545
https://doi.org/10.1140/epjp/s13360-022-02733-4

THE EUROPEAN PHYSICAL JOURNAL PLUS

Regular Article

Analysis of melting of some metals using pressure dependence of the Grüneisen parameter in the Lindemann law

D. Ashwini[1,a], **Vijay S. Sharma**[2,b], **K. Sunil**[3,c]

[1] Department of Physics, I. B. S. Khandari Campus, Dr. Bhimrao Ambedkar University, Agra 282002, India
[2] Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India
[3] Department of Physics, I. E. T. Khandari Campus, Dr. Bhimrao Ambedkar University, Agra 282002, India

Received: 21 January 2022 / Accepted: 17 April 2022
© The Author(s), under exclusive licence to Società Italiana di Fisica and Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract In order to obtain the melting curves representing the relationship between melting temperature and pressure for a material with the help of Lindemann's law, we need an accurate evaluation of the bulk modulus and the Grüneisen parameter at different pressures. We have used the results determined from the reciprocal K-primed equation of state and the reciprocal gamma relationship in the Lindemann formulation to calculate melting temperatures and slopes of the melting curves for ten metals, viz. Cu, Mg, Pb, Al, In, Cd, Zn, Au, Ag, Mn at different pressures. The results are found to present reasonably good agreement with the experimental data up to 16 GPa, except in few cases such as Al. The melting curves of five metals, viz. Cu, Mg, Zn, Au, Ag, are extended into a megabar pressure range using the present method and compared with the data reported in the recent literature. The agreement with the experimental data is improved to a great extent when we use the Stacey law of melting in place of the Lindemann law. The method developed in the present study has the advantage that the melting curve can be determined analytically without using power computers.

1 Introduction

Arafin et al. [1, 2] have presented a formulation for the pressure dependence of melting temperature T_m for a material which yields the following expression

$$T_m(P) = T_{m0}\left(1 + \frac{K_0' P}{K_0}\right)^{2(\gamma_0 - 1/3)/K_0'} \tag{1}$$

where T_{m0} is the melting temperature, γ_0 is the Grüneisen parameter, K_0 is the bulk modulus, K_0' is the pressure derivative of bulk modulus, all at zero pressure (P = 0). Equation (1) has been derived [1, 2] from the Lindemann law [3, 4] incorporating two assumptions. First, bulk modulus of solids depends linearly on pressure

$$K = K_0 + K_0' P \tag{2}$$

and the second assumption is that the Grüneisen parameter does not depend on pressure, i.e., $\gamma(P) = \gamma_0$. Both of these assumptions are not valid. Variation of K with P does not follow Eq. (2) because the bulk modulus depends nonlinearly on pressure such that the slope dK/dP decreases with the increase in pressure [5] and attains a constant values equal to K_∞' in the limit of infinite pressure [6]. The Grüneisen parameter γ is not a constant quantity. The pressure derivatives of gamma have very significant magnitudes in terms of second-order and third-order Grüneisen parameters, q and λ, respectively. They are useful for investigating higher-order thermoelastic properties of solids [7, 8]. In spite of the invalidity of the two assumptions, Eq. (1) has been found to yield reasonable agreement with the experimental melting data [9, 10]. It appears therefore that there is a lucky cancellation of errors arising from the two inaccurate assumptions employed in the derivation of Eq. (1) by Arafin et al. [1, 2].

Subsequently Arafin and Singh [11, 12] considered the following quadratic expressions for bulk modulus K and the Grüneisen parameter γ in powers of pressure

$$K = K_0 + a_1 P + a_2 P^2 \tag{3}$$

[a] e-mail: adashwinidixit@gmail.com
[b] e-mail: vss7694@gmail.com
[c] e-mail: k.sunil.ibs@gmail.com (corresponding author)

Published online: 04 May 2022

Table 1 Values of input parameters of metals K_0, K_0' and γ_0 from the compilation by Kholiya and Chandra [9]. The other parameters K_∞', γ_∞, A and B are calculated as described in the text

Metal	K_0 (GPa)	K_0'	K_∞'	γ_0	γ_∞	A Equation (11)	B Equation (12)
Cu	133.5	3.88	2.33	2.01	1.00	0.4975	1.1708
Mg	36.8	4.30	2.58	1.51	1.12	0.6623	0.5949
Pb	42.7	5.70	3.42	2.63	1.54	0.3802	0.9204
Al	73	4.50	2.70	2.17	1.18	0.4608	1.0439
In	42.68	480	2.88	2.04	1.27	0.4902	0.8560
Cd	42	6.50	3.90	2.40	1.78	0.4167	0.5660
Zn	65	4.60	2.76	2.25	1.21	0.4444	1.0543
Au	167	5.70	3.42	3.03	1.54	0.3300	1.0921
Ag	103.6	4.70	2.82	2.40	1.24	0.4167	1.0992
Mn	158	4.60	2.76	1.95	1.21	0.5128	0.8656

and

$$\gamma = \gamma_0 + b_1 P + b_2 P^2 \tag{4}$$

where K_0 and γ_0 are the values of K and γ, respectively, both at zero pressure, a_1, a_2, b_1 and b_2 are constants for a given solid. Using Eqs. (3) and (4) in the Lindemann law of melting

$$\frac{d \ln T_m}{dP} = \frac{2\left(\gamma - \frac{1}{3}\right)}{K} \tag{5}$$

and then integrating Eq. (5), Arafin and Singh [11, 12] calculated values of T_m at different pressures for metals as well as non-metals. However, it has been found [13, 14] that Eqs. (3) and (4) are not valid at high pressures. Hence the model used by Arafin and Singh [11, 12] is not physically acceptable. Equations (3) and (4) yield the Grüneisen parameter, bulk modulus and its pressure derivative which are not consistent with the thermodynamic constraints [5, 6]. Values of γ and K' cannot be negative for normal materials at any pressure [15, 16]. From the point of view of melting, normal materials are those which do not exhibit any maxima or minima in their melting curves.

Recently Sunil et al. [17] have developed a method for determining melting curves of MgO and LiF using the Lindemann law by taking into account the pressure dependence of Grüneisen parameter with the help of the reciprocal gamma relationship in terms of *P/K* [15]. This relationship has been found to be consistent with the Stacey reciprocal *K*-primed equation of state [18] and also with the thermodynamic constraints for the higher-order thermoelastic properties in the limit of infinite pressure [15]. In view of the successful application of the model in case of MgO and LiF [17], we apply it to calculate melting curves of ten metals viz. Cu, Mg, Pb, Al, In, Cd, Zn, Au, Ag, Mn. The results are compared with the available experimental data [19–28] and also with those obtained from a theoretical model [9] based on the instability of the thermal equation of state originally due to Shanker et al. [29, 30]. We make use of the reliable values of P, K and K' for the metals under study determined with the help of the Stacey reciprocal K-primed equation of state [5, 8, 18]. The relevance of the present study becomes clear from the recent important investigations based on the experimental techniques using the Laser-heated diamond anvil cells and synchrotron X-ray diffraction measurements, and based on the ab initio simulation method to determine melting curve and phase diagram of Mg [31], Zn [32], Pt [33], Nb [34], Au [35], Ir [36], and Cu [37]. The theoretical methods include the accurate DFT calculations and molecular dynamics which are very time-consuming and have high cost of computing. On the other hand, the present method based on the Lindemann law can be used conveniently for determining melting curves without the help of power computers.

Hieu and Ha [38] obtained high-pressure melting curves of silver, gold and copper using the volume dependence of gamma, $\gamma = \gamma_0(V/V_0)^q$ with a constant value of q equal to one. Further, they used the Vinet equation of state [39] to transform the volume data to corresponding values of pressure in order to determine melting temperatures at different pressures. It has been discussed at length by Stacey and Davis [5, 6] that the volume dependence of γ with $q = 1$ and also the Vinet EOS are not consistent with the thermodynamic constraints for solids. In the present study, we have used valid expressions to change gamma in terms of K and P. This is an innovation in the calculation method. An additional innovation for calculating the melting temperatures under different pressures is the procedure based on Euler's finite difference calculus.

It is worth mentioning here that Ma et al. [40] have recently developed a model for the pressure dependence of melting temperature for metals based on the Force-Heat Equivalence Energy Density principle [41]. An equivalent relationship between the heat energy variations during cooling and the mechanical work during pressurization for the solidification of molten metal has been applied to derive an expression for melting temperature at different pressures in terms of bulk modulus K and its first pressure derivative K', both at zero pressure [40]. The model thus developed by Ma et al. predicts melting curves for metals in good agreement with the

Table 2 Values of V/V_0, P, K, K' determined from the Stacey reciprocal equation of state

Cu				Mg			
V/V_0	P (GPa)	K (GPa)	K'	V/V_0	P (GPa)	K (GPa)	K'
1	0	133.5	3.88	1	0	36.8	4.30
0.9946	0.735	136.3	3.85	0.9809	0.74	39.9	4.17
0.9877	1.7	140	3.81	0.9580	1.73	44	4.03
0.9792	2.93	144.7	3.76	0.9482	2.19	45.8	3.97
0.9743	3.66	147.4	3.74	0.9288	3.18	49.7	3.87
0.9720	4.0	148.7	3.72	0.9113	4.15	53.3	3.79
0.9678	4.65	151.1	3.70	0.8957	5.12	57.1	3.72
0.9617	5.63	154.8	3.67	0.8740	6.58	62.4	3.64
0.9539	6.9	159.4	3.64	0.8566	7.81	66.5	3.58
0.9373	9.8	169.8	3.56	0.8470	8.54	69	3.54
0.9214	12.8	180.4	3.50	0.8326	9.75	73.2	3.50
0.9056	16	191.4	3.44	0.8188	10.97	77.3	3.46
Pb				Al			
V/V_0	P (GPa)	K (GPa)	K'	V/V_0	P (GPa)	K (GPa)	K'
1	0	42.7	5.70	1	0	73	4.50
0.9829	0.77	47	5.49	0.9899	0.757	76.4	4.42
0.9683	1.5	50.9	5.34	0.9781	1.7	80.5	4.33
0.9470	2.7	57.1	5.14	0.9629	3	86	4.23
0.9390	3.2	59.7	5.08	0.9478	4.4	91.9	4.14
0.9238	4.2	64.6	4.97	0.9399	5.2	95.3	4.10
0.9102	5.2	69.5	4.87	0.9256	6.7	101.4	4.02
0.8952	6.4	75.3	4.78	0.9126	8.2	107.5	3.96
0.8735	8.1	81.9	4.65	0.8987	9.9	114.2	3.89
0.8567	9.6	88.1	4.57				
0.8406	11.1	93.9	4.49				
0.8330	11.9	97.2	4.46				
In				Cd			
V/V_0	P (GPa)	K (GPa)	K'	V/V_0	P (GPa)	K (GPa)	K'
1	0	42.68	4.80	1	0	42	6.50
0.9841	0.71	46	4.66	0.9844	0.69	46.4	6.26
0.9549	2.2	52.8	4.44	0.9740	1.2	49.6	6.11
0.9410	3	56.4	4.36	0.9076	5.5	74.4	5.45
0.9178	4.5	62.9	4.22	0.8939	6.7	80.8	5.34
0.9120	4.9	64.6	4.19	0.8834	7.7	85.9	5.27
0.8973	6.0	69.2	4.12	0.8693	9.2	93.5	5.18
0.8817	7.2	73.8	4.04	0.8573	10.7	100.5	5.11
0.8763	7.7	76.0	4.02	0.8485	11.9	105.9	5.06
0.8590	9.2	81.6	3.95				
0.8385	11.2	89.2	3.87				
0.8308	12	92.1	3.84				

Table 2 continued

Zn				Au			
V/V_0	P (GPa)	K (GPa)	K'	V/V_0	P (GPa)	K (GPa)	K'
1	0	65	4.60	1	0	167	5.70
0.9904	0.64	67.9	4.52	0.9975	0.42	169.4	5.67
0.9838	1.1	70	4.47	0.9930	1.2	173.8	5.61
0.9576	3.1	78 .7	4.29	0.9874	2.2	179.4	5.54
0.9381	4.8	85.9	4.17	0.9820	3.2	184.9	5.48
0.9297	5.6	89.3	4.12	0.9794	3.7	187.6	5.46
0.9177	6.8	94.3	4.06	0.9742	4.7	193.1	5.40
0.9090	7.7	97.8	4.02	0.9698	5.6	198.0	5.36
0.8942	9.4	104.9	3.95	0.9640	6.8	204.4	5.30
0.8855	10.4	108.6	3.91				
0.8758	11.6	113.1	3.87				
Ag				Mn			
V/V_0	P (GPa)	K (GPa)	K'	V/V_0	P (GPa)	K (GPa)	K'
1	0	103.6	4.70	1	0	158	4.60
0.9930	0.74	107.1	4.64	0.9951	0.76	161.5	4.56
0.9800	2.2	113.7	4.54	0.9907	1.5	164.8	4.52
0.9716	3.2	118.3	4.47	0.9878	2.0	167.1	4.50
0.9674	3.7	120.5	4.44	0.9836	2.7	170.3	4.47
0.9596	4.7	124.9	4.39	0.9808	3.2	172.5	4.45
0.9544	5.4	128.0	4.36				
0.9499	6.0	130.6	4.33				
0.9432	6.9	134.5	4.29				
0.9387	7.6	137.5	4.26				

available experimental data [27] and also provides insights into the fundamental understanding of quantitative effect of pressure on melting temperature.

2 Method of analysis

In the limit of extreme compression, volume V tends to zero, pressure P tends to infinity, K also tends to infinity, but the ratio P/K remains finite [42]

$$\left(\frac{P}{K}\right)_{\infty} = \frac{1}{K'_{\infty}} \tag{6}$$

Equation (6) has the status of an identity. The subscript ∞ represents the limit of infinite pressure. K'_{∞} is the value of $K' = dK/dP$ in the limit of infinite pressure. K' decreases with the increase in pressure and attains a minimum value equal to K'_{∞} in the limit of infinite pressure. The theory of infinite pressure extrapolation has been developed [5, 6] by considering a material to remain in the same structure or phase, i.e., the material is supposed to be compressed down to extreme limit, volume tends to zero without exhibiting any phase transitions. Using Eq. (6), Stacey [18] proposed the following relationship for reciprocal K'

$$\frac{1}{K'} = \frac{1}{K'_0} + \left(1 - \frac{K'_{\infty}}{K'_0}\right)\frac{P}{K} \tag{7}$$

On integrating Eq. (7), we get

$$\frac{K}{K_0} = \left(1 - \frac{K'_{\infty} P}{K}\right)^{-K'_0/K'_{\infty}} \tag{8}$$

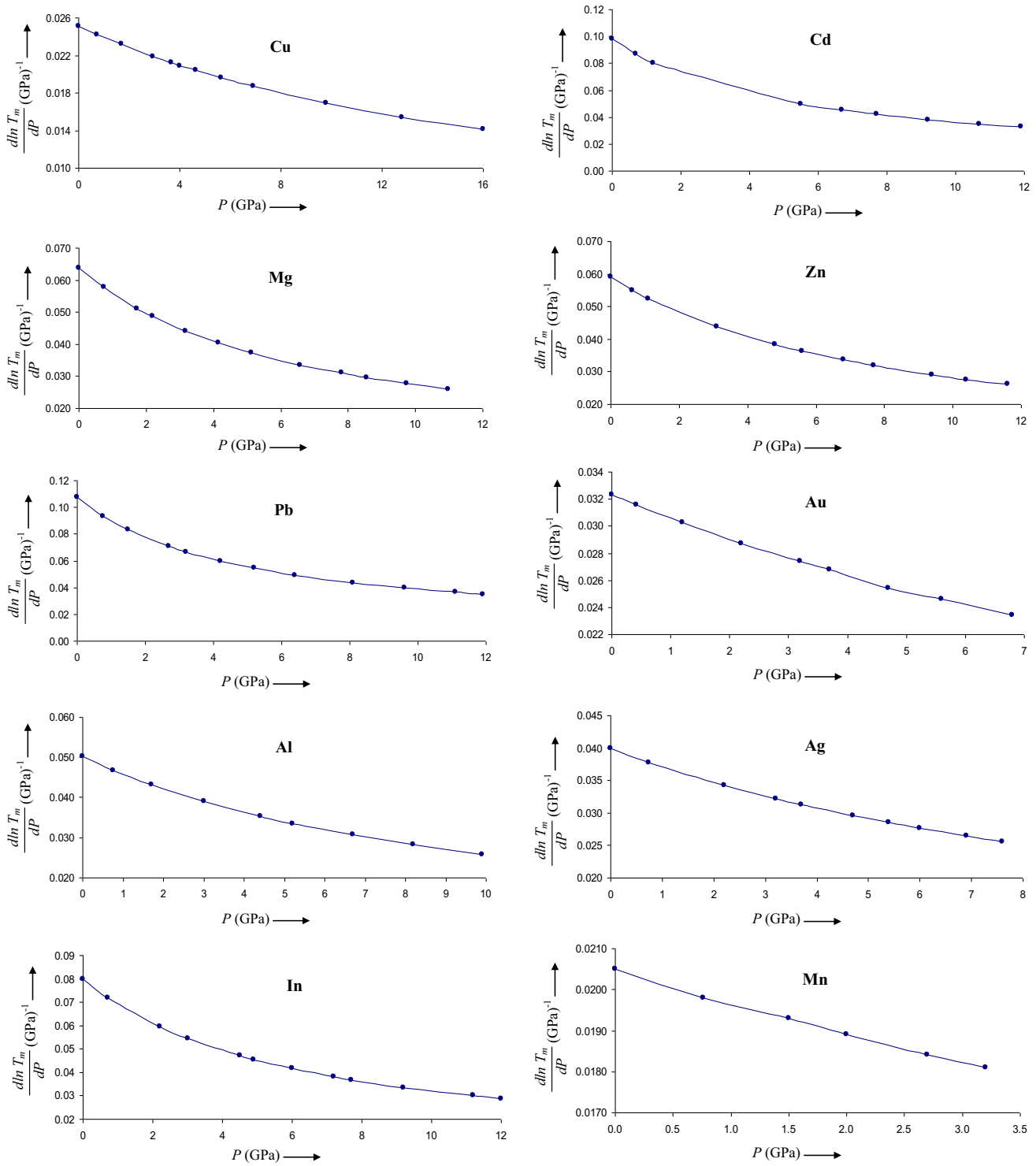


Fig. 1 Melting slopes for ten metals calculated from Eqs. (13) and (15)

Further integration of Eq. (8) gives

$$\ln\left(\frac{V}{V_0}\right) = \left(\frac{K_0'}{K_\infty'} - 1\right)\frac{P}{K} + \frac{K_0'}{K_\infty'^2}\ln\left(1 - K_\infty'\frac{P}{K}\right) \tag{9}$$

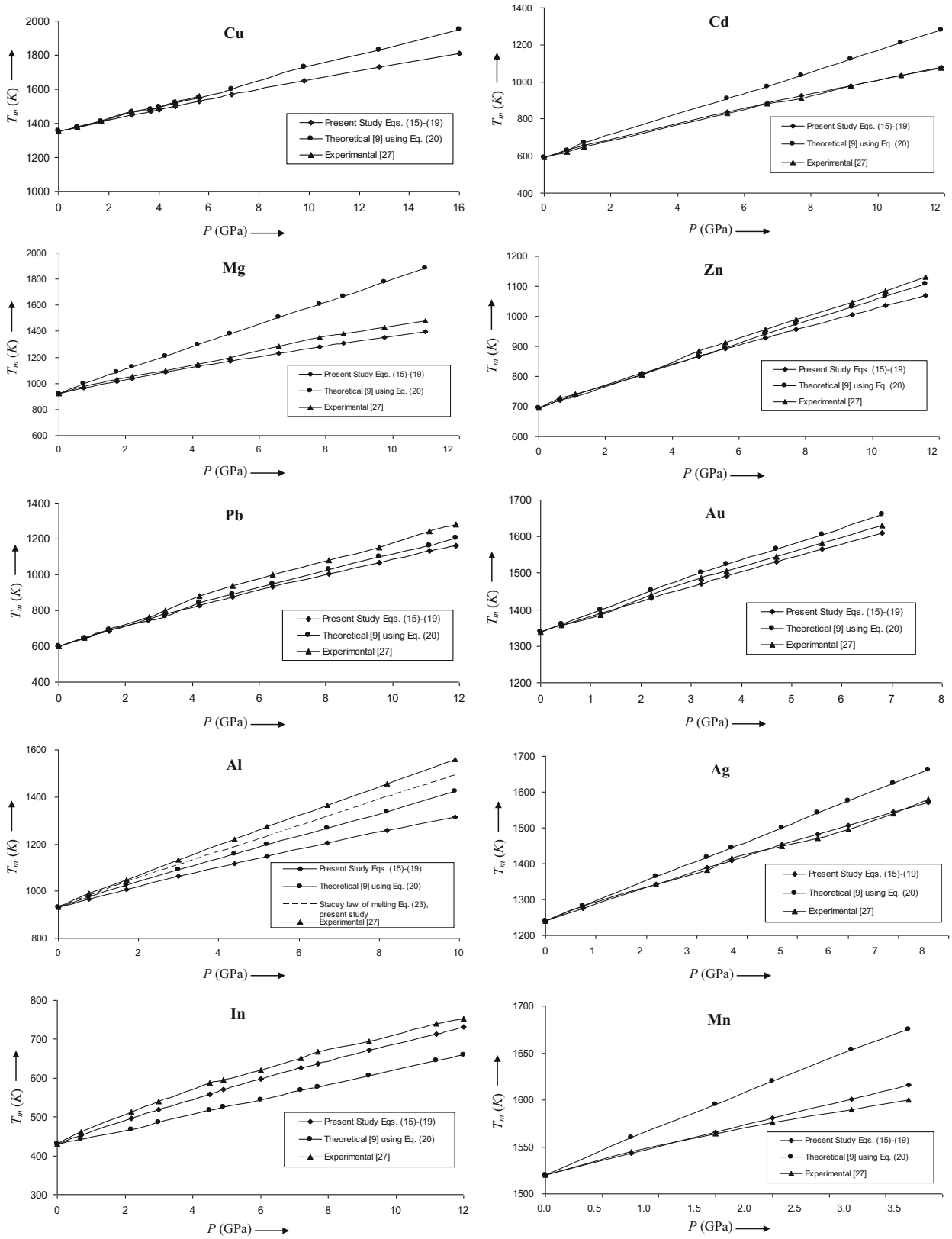


Fig. 2 Melting curves for ten metals

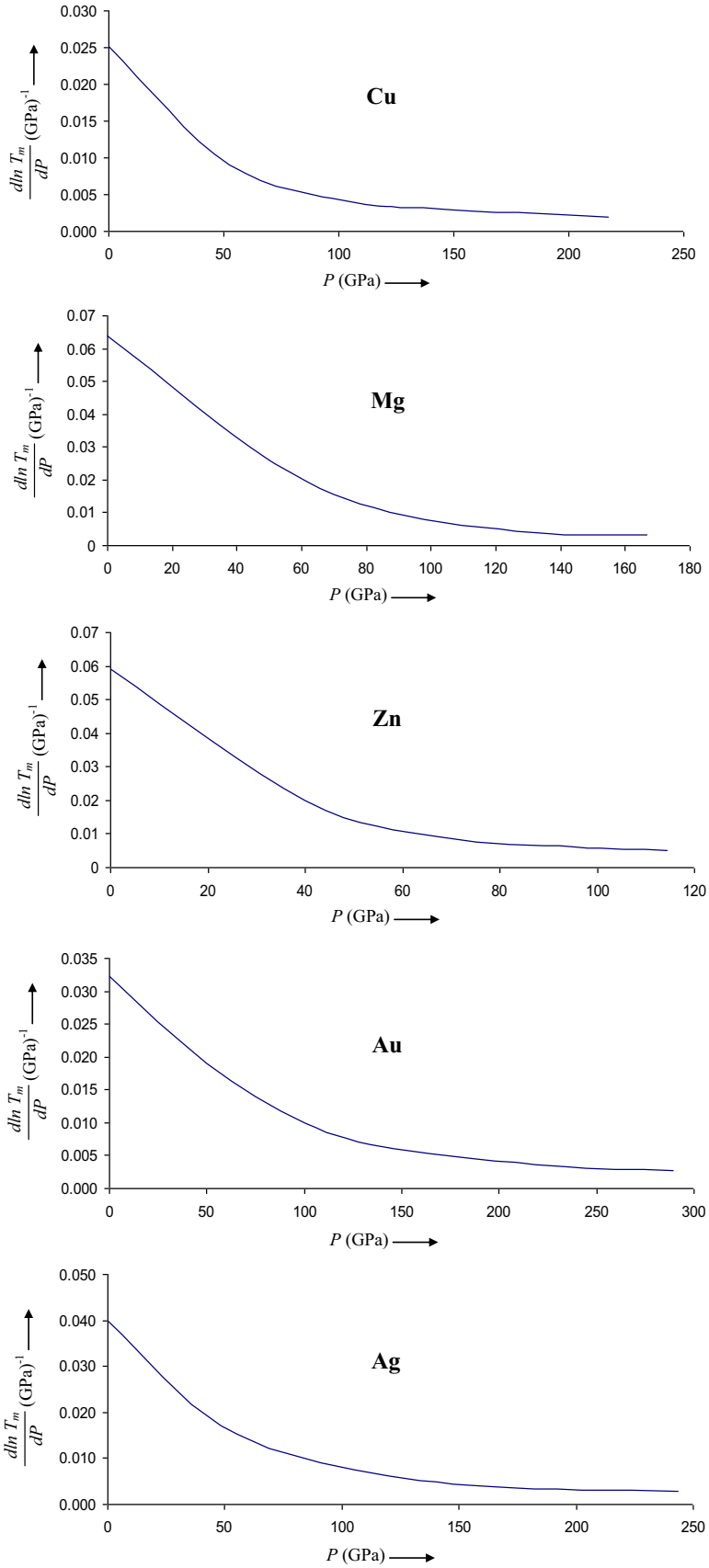


Fig. 3 Melting slopes for five metals at high pressures calculated from Eqs. (13) and (15)

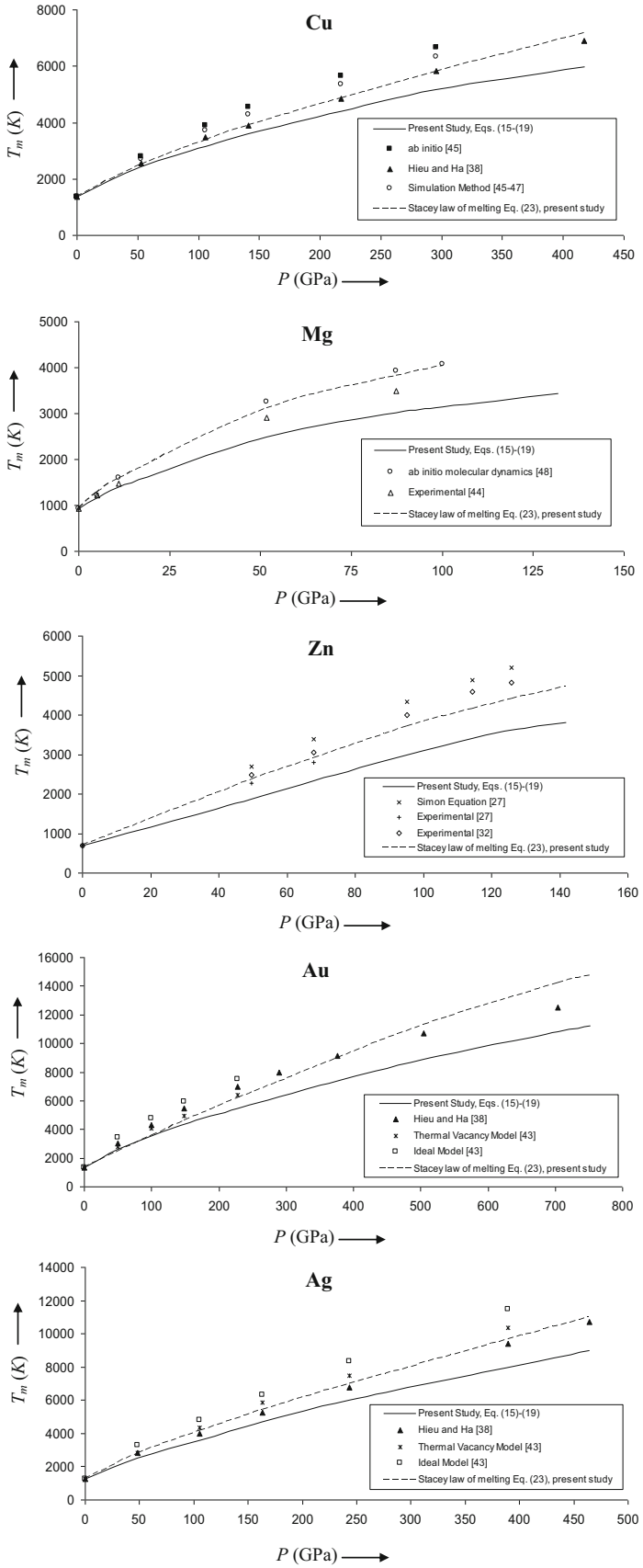


Fig. 4 Melting curves for some metals at high pressures

where V_0 is the value of volume V at $P = 0$. Realizing that γ and K' change with pressure in a similar manner, both decrease with the increase in pressure, and attain constant minimum values γ_∞ and K'_∞, respectively, Shanker et al. [15] proposed the reciprocal gamma relationship analogous to Eq. (7) as follows:

$$\frac{1}{\gamma} = A + B\frac{P}{K} \tag{10}$$

where

$$A = \frac{1}{\gamma_0} \tag{11}$$

and

$$B = K'_\infty\left(\frac{1}{\gamma_\infty} - \frac{1}{\gamma_0}\right) \tag{12}$$

Using Eq. (10) in the Lindemann law, Eq. (5), we get

$$\frac{d\ln T_m}{dP} = \frac{2}{AK + BP} - \frac{2}{3K} \tag{13}$$

The input parameters used in calculations are given in Table 1. Values of K_0, K'_0 and γ_0 are based on the experimental data compiled by Kholiya and Chandra [9]. Values of K'_∞ are determined using the relationship $\mathrm{K}'_\infty = (3/5)\mathrm{K}'_0$ given by Stacey [6]. The validity of this relationship has been discussed at length in an earlier paper [17]. Values of γ_∞ are then calculated from the following equation [5, 6]

$$\gamma_\infty = \frac{1}{2}K'_\infty - \frac{1}{6} \tag{14}$$

Equation (14) has the status of an identity. The results for V/V_0, P, K, K' are given in Table 2 obtained using the input parameters in Table 1 and the Stacey equations (7)–(9).

3 Results and discussion

Using the method based on Euler's finite difference calculus, Eq. (13) can be rewritten as follows:

$$\frac{1}{T_m}\frac{\Delta T_m}{\Delta P} = \Phi \tag{15}$$

where

$$\Phi = \frac{2}{AK + BP} - \frac{2}{3K} \tag{16}$$

and, then we get

$$\begin{aligned}\Delta T_m &= T_m - T_{m0}\\ &= \Phi\Delta P\, T_m\end{aligned} \tag{17}$$

so that

$$T_m = \frac{T_{m0}}{1 - \Phi\,\Delta P} \tag{18}$$

A stepwise calculation can be performed, for example, by taking ΔP = 1GPa . T_m at 1 *GPa* is determined by taking Φ at 1 *GPa* and the experimental value of T_{m0} at zero pressure. T_m at 1 *GPa* is then designated by T_{m1} and T_m at 2 *GPa* by Tm_2 such that

$$T_{m2} = \frac{T_{m1}}{1 - \Phi\Delta P} \tag{19}$$

Values of T_{m1} at 1 *GPa* have been used to determine T_{m2} at 2 *GPa* taking Φ at 2 *GPa,* and $\Delta P = 1GPa$. We proceed further in a similar manner to determine T_m at higher pressures. The accuracy of Euler's method is increased when the size of interval ΔP is reduced. Calculations can be performed for any finite value of interval ΔP. To determine T_{m2} at any pressure P, we have to use T_{m1} at pressure $P - \Delta P$, and Φ at the same pressure P. In fact, Φ represents the melting slope $dlnT_m/dP$ (Eqs. 13 and 16). The results for Φ determined from Eq. (16) are shown in Fig. 1 for the ten metals under study. Results for the melting curves are given

in Fig. 2 and compared with the available experimental data [19–28]. Kholiya and Chandra [9] have used an alternative method for determining melting curves based on the instability of thermal equation of state [29, 30] which yields

$$\frac{T_m - T_0}{T_{m0} - T_0} = \frac{K\left(K_0' + 1\right)}{K_0(K' + 1)} \tag{20}$$

where T_0 is the reference temperature or room temperature. Values of K_0 and K_0' are taken from Table 1. K and K' at different pressures are given in Table 2. The melting curves were determined by Kholiya and Chandra [9] using Eq. (20). These are also included in Fig. 2 for the sake of comparison. The two theoretical models Eq. (13) and Eq. (20), yield results for melting curves in reasonably good agreement with each other, and also with the experimental data [19–28]. There are few exceptions such as Al for which the calculated values of T_m differ substantially from the experimental data (Fig. 2). This is possibly due to the inadequacy of the Lindemann law in case of some metals [14]. Using some basic principles of thermodynamics of solids, Stacey obtained the following formulation (Eq. (80) of Reference [6])

$$\frac{d \ln T_m}{d \ln V} = -2\gamma \tag{21}$$

which gives

$$\frac{1}{T_m}\left(\frac{dT_m}{dP}\right) = \frac{2\gamma}{K} \tag{22}$$

where $K = -V(dP/dV)$. With the help of Eq. (10) in Eq. (22), we get

$$\frac{1}{T_m}\left(\frac{dT_m}{dP}\right) = \frac{2}{AK + BP} \tag{23}$$

High-pressure melting curves can be determined with the help of Eq. (23) using the Euler finite difference method. The results for Al thus obtained and shown in Fig. 2 are in better agreement with the experimental data.

The melting curves of five metals, viz. Cu, Mg, Zn, Au, Ag, are extended into a megabar pressure range using the Lindemann law. Calculated values of melting slopes are given in Fig. 3 and melting curves in Fig. 4. A comparison is presented in Fig. 4 with the experimental data [27, 32, 38, 43, 44] and with the ab initio molecular dynamics simulations [45–48]. It is found that the values determined from the Lindemann law shown by continuous curves in Fig. 4 are significantly lower than the experimental data and simulation studies. When we use the Stacey law of melting (Eqs. 21–23) in place of the Lindemann law, the agreement between calculated and experimental values is improved to a great extent.

4 Conclusions

The earlier models developed by Arafin et al. [1, 2, 11, 12] are not adequate for describing high pressure melting of solids because of invalid assumptions regarding the pressure dependences of bulk modulus and the Grüneisen parameter [13, 14]. It is found here that for developing an adequate theory of high-pressure melting, we should take an appropriate account of the volume dependence or pressure dependence of gamma in the Lindemann law. This has been done in the present study by using the relationship [15] between reciprocal gamma and *P/K* which is consistent with the Stacey thermodynamic constraints. The melting temperatures for the metals at high pressures calculated from the Lindemann law are somewhat lower than the available experimental data. The agreement is much improved at high pressures when we use the Stacey law of melting in place of the Lindemann law.

A remarkable feature of the method based on Eq. (13) is that it directly yields values of melting slopes ($dlnT_m/dP$) in terms of P and K in the denominator. The melting slope is therefore found to decrease regularly with the increase in pressure and tending to zero in the limit of infinite pressure. This finding is supported by recent studies in the field of melting at high pressures [9, 14, 49]. The melting curves obtained from the vacancy model [49] are of similar nature as those based on the Lindemann law. This is also confirmed from the recent work on the thermal vacancy model by Cuong and Phan [43]. The Lindemann law as well as the thermal vacancy model, both are based on the thermodynamic instability of the solid phase.

Acknowledgements We are thankful to the reviewer for his helpful comments which have been used in revising the manuscript. Authors are also thankful to Professor Jai Shanker for some useful suggestions.

Data availability "This manuscript has associated data in a data repository. [Authors' comment: The data that support the finding of this study are available within the article.]"

References

1. S. Arafin, R.N. Singh, A.K. George, Int. J. Thermophyics **33**, 1013 (2012)
2. S. Arafin, R.N. Singh, A.K. George, Physica B **419**, 40 (2013)

3. F.A. Lindemann, Z. Phys. **11**, 609 (1910)
4. J.J. Gilvarry, Phys. Rev. **102**, 307 (1956)
5. F.D. Stacey, P.M. Davis, Phys. Earth Planet. Int. **142**, 137 (2004)
6. F.D. Stacey, Rep. Prog. Phys. **68**, 341 (2005)
7. J. Shanker, P. Dulari, P.K. Singh, Physica B **404**, 4083 (2009)
8. J. Shanker, K. Sunil, B.S. Sharma, Physica B **407**, 2082 (2012)
9. K. Kholiya, J. Chandra, Modern Phys. Lett. B **29**, 1550161 (2015)
10. M. Goyal, Chin. J. of Phys. **66**, 453 (2020)
11. S. Arafin, R.N. Singh, Int. J. Mod. Phys. B **30**, 1750031 (2016)
12. S. Arafin, R.N. Singh, J. Phys. Chem. Solids **91**, 101 (2016)
13. A. Vijay, Int. J. Mod. Phys. B **33**, 1975001 (2019)
14. J. Shanker, K. Anand, B.S. Sharma, A. Vijay, Int. J. Thermophysics **41**, 1 (2020)
15. J. Shanker, K. Sunil, B.S. Sharma, Phys. Earth Planet. Int. **262**, 41 (2017)
16. F.D. Stacey, J.H. Hodgkinson, Phys. Earth Planet. Int. **286**, 42 (2019)
17. K. Sunil, S.B. Sharma, B.S. Sharma, Inter. J. Modern Phys. B **30**, 1850339 (2018)
18. F.D. Stacey, Geophys. J. Int. **143**, 621 (2000)
19. N.R. Mitra, D.L. Decker, H.B. Vanfleet, Phys. Rev. **161**, 613 (1967)
20. J. Akella, G.C. Kennedy, J. Geophysical, Research **76**, 4969 (1971)
21. J.F. Cannon, J. Phys. Chem. Ref. Data **3**, 781 (1974)
22. J.A. Moriarty, J.D. Althoff, Phys. Rev. B **51**, 5609 (1995)
23. G. Shen, N. Sata, M.L. Rivers, S.R. Sutton, Appl. Phys. Lett. **78**, 3208 (2001)
24. M. Minicucci, A. Trapananti, A.D. Cicco, S.D. Panfilis, G. Aquilanti, Phys. Scr. T. **115**, 1056 (2005)
25. H. Brand, D.P. Dobson, L. Vocadlo, I.G. Wood, High Pressure Res. **26**, 185 (2006)
26. F. Cricchio, A.B. Belonoshko, L. Burakovsky, D.L. Preston, R. Ahuja, Phys. Rev. B **73**, 140103 (2006)
27. D. Errandonea, J. Appl. Phys. **108**, 033517 (2010)
28. P.D. Tam, P.D. Tan, N.Q. Hoc, P.D. Phong, Proc. Natl. Conf. Theor. Phys. **35**, 148 (2018)
29. J. Shanker, S.S. Kushwah, P. Kumar, Physica B **239**, 337 (1997)
30. J. Shanker, M.P. Sharma, S.S. Kushwah, J. Phys. Chem. Solids **60**, 603 (1999)
31. G.W. Stinton, S.G. MacLeod, H. Cynn, D. Errandonea, W.J. Evans, J.E. Proctor Meng et al., Phys. Rev. B **90**, 134105 (2014)
32. D. Errandonea, S.G. MacLeod, J. Ruiz-Fuertes, L. Burakovsky, M.I. McMahon, C.W. Wilson et al., J. Phys. Condens. Matter **30**, 295402 (2018)
33. S. Anzellini, V. Monteseguro, E. Bandiello, A. Dewaele, L. Burakovsky, D. Errandonea, Sci. Rep. **9**, 13034 (2019)
34. D. Errandonea, L. Burakovsky, D.L. Preston, S.G. MacLeod, D. Santamaria-Perez, S. Chen et al., Commun. Mater. **1**, 60 (2020)
35. G. Weck, V. Recoules, J.-A. Queyroux, F. Datchi, J. Bouchet, S. Ninet et al., Phys. Rev. B **101**, 014106 (2020)
36. S. Anzellini, L. Burakovsky, R. Tumbull, E. Bandiello, D. Errandonea, Crystals **11**, 452 (2021)
37. S.R. Baty, L. Burakovsky, D. Errandonea, Crystals **11**, 537 (2021)
38. H.K. Hieu, N.N. Ha, AIP Adv. **3**, 112125 (2013)
39. P. Vinet, J. Ferrante, J. Rose, J. Smithj, J. Geophy. Res. **92**, 9319 (1987)
40. J. Ma, W. Li, G. Yang et al., Phys. Earth Planet. Inter. **309**, 106602 (2020)
41. W.G. Li, F. Yang, D.N. Fang, Acta Mech. Sinica **26**, 235 (2010)
42. L. Knopoff, J. Geophys. Res. **68**, 2932 (1963)
43. T.D. Cuong, A.D. Phan, Vacuum **179**, 109444 (2020)
44. D. Errandonea, R. Boehler, M. Ross, Phys. Rev. B **65**, 012108 (2001)
45. Y.N. Wu, L. Wang, Y. Huang, D. Wang, Chem. Phys. Lett. **515**, 217 (2011)
46. S. Wang, G. Zhang, H. Liu, H. Song, J. Chem. Phys. **138**, 134101 (2013)
47. Z.L. Liu, X.L. Zhang, L.C. Cai, J. Chem. Phys. **143**, 114101 (2015)
48. S. Mehta, G.D. Price, D. Alfe, J. Chem. Phys. **125**, 194507 (2006)
49. K. Ksiazek, T. Gorecki, High Temp.-High Press. **32**, 185 (2020)

Journal of Ambient Intelligence and Humanized Computing
https://doi.org/10.1007/s12652-022-03878-x

ORIGINAL RESEARCH

Analysis of wind turbine dataset and machine learning based forecasting in SCADA-system

Upma Singh[1,2] · **M. Rizwan**[1]

Received: 25 January 2022 / Accepted: 27 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
In this paper, Machine Learning (ML) based techniques known as Support Vector Regression (SVR) and Gradient Boosting Regression Trees (GBRT) are utilized for improving the forecast accuracy of short-term wind power production. In addition, data analysis framework is also prepared to visualize the data acquired from the Supervisory Control and Data Acquisition (SCADA) system. Dataset analysis has been carried out in Polar and Cartesian coordinates to identify the produced power and wind relationship. Based upon different statistical indices, the performance of the developed model has been evaluated. In addition, the comparative analysis reveals that GBRT model performance is superior than that of the SVR model with Mean-Square Error (MSE) value of 0.026, Root Mean-Square Error (RMSE) 0.301, Mean Absolute Percentage Error (MAPE) 0.063, Mean Absolute Error (MAE) 0.004 and Coefficient of determination (R^2) 0.969. The results obtained illustrate that the proposed model would be applicable in real-life scenarios and help the management team to handle the produced power from wind turbines.

Keywords Regression techniques · Forecasting models · Machine learning · Wind power generation · Renewable energy sources

1 Introduction

With the advent of new technologies and accelerated growth in the world's economy, the power demand increases substantially. Over the recent times, power industries have switched their focus on to the sources of green power in order to minimize the carbon emission during power production (Ahmad et al. 2021; Yuan et al. 2021). In addition, fossil fuels has a severe influence on the environment such as floods, melting of glaciers, heat waves, droughts and frequent wildfires threatening our earth's ecosystem. Thus, the growth in renewable energy sources (RES) like solar power, geothermal energy, tidal energy and wind energy have become the long-term goal for the governments throughout the world (Deng et al. 2020). Among all power sources, the wind is the source of green power which can replace the utilization of fossil fuels, thereby reducing emission of carbon and efficiently lessening the power crisis (Qin et al. 2021a, b; Wang et al. 2020). Wind has broadly been adopted as power-system alternative because of its easy accessibility, cost-effectiveness and renewable nature. Because of the intermittent nature of power production from wind energy systems, the operators have to operate and control the power-plant competently (Bo et al. 2021; Jacondino et al. 2021). Therefore, for long-term and short-term planning of power transmission, it is necessary to forecast wind power production. Recently numerous researches have been done on predicting the power production and feasibility analysis by wind system. Few researches have concentrated on direct-power forecasting of RES and some have utilized the methodology to predict the input-variables (wind and solar-irradiance) and estimated the energy produced from hybrid plant utilizing various models (Decaro et al. 2021). Mostly wind-farms have installed SCADA system to manage the wind turbines for logging on the operational data on and monitoring distinct components. The logged on dataset

✉ Upma Singh
upma14@gmail.com

M. Rizwan
rizwan@dce.ac.in

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

[2] Department of Electronics and Communication Engineering, Maharaja Surajmal Institute of Technology, Delhi 110058, India

Published online: 27 May 2022

contains information about speed of wind, direction of wind and produced power etc. It provides an opportunity to process the gathered time-series data for diverse applications ranging from operational and maintenance purposes to predictive analyses of produced power via wind turbines (Hu et al. 2021). The wind speed, wind direction and time of day are the significant parameters which influence the wind turbine-power production. In the past few years, numerous research papers and review articles have been published that analyzed, in detail, the various wind power forecasting approaches and methods. Hanifi et al. (2020) presented a critical review on various approaches such as statistical, physical and hybrid to forecast the wind power. In this area of research, authors also provide the past and present trends and highlight the future research directions that incorporate a requirement to evolve more cost-effective and advanced forecasting approaches; improvement in data-processing, error post processing and developing specific models for an off-shore wind energy forecasting, as such wind turbines work in distinct weather circumstances. Moreover, their findings show that, among 42 analyzed researches 60% utilized wind speed, 25% utilized air temperature and few researches utilized number of generation hours.

Delgado and Fahim (2021) evolved long-short term memory (LSTM) based forecasting model for short-term wind energy production. In another research, Extra Trees, ADaBoost and K-Neighbors Regressors are used to predict power and energy of wind turbine. While manipulations of data are accomplished by recursive-feature-elimination utilizing cross-validation (RFECV) (Qadir et al. 2021) In another research, a new forecast model using light gradient boosting machine (GBM) and convolutional neural network (CNN) is developed. By examining the characteristics of raw-dataset on the time-series from the surrounding wind field, new feature sets are built, and CNN are utilized to recover the details from the input dataset, and network parameters are updated by comparing the outputs Ju et al (2019). Chandran et al. (2021) evolved a model that could effectively predict wind energy by using machine algorithms such as recurrent neural network (RNN), long short-term Memory (LSTM) and gated reference unit (GRU). In other work, to predict the next hour wind power a sparse machine learning technique is utilized (Lv et al. 2021). Kisvari et al. (2021) evolved a novel data-driven methodology to predict wind energy by incorporating dataset preprocessing, detection of anomalies and treatment, hyper-parameter tuning and feature engineering using gated-recurrent deep-learning models with the help of six features such as gearbox temperature, generator temperature, nacelle orientation, wind speed at hub height, blade pitch angle and rotor speed. In another work, comparison of five optimized robust regression models is done to predict wind turbine output on the basis of wind velocity vector components (Pathak et al. 2021). González-Sopeña et al. (2021) presented a widespread survey of the performance assessment approaches utilized for evaluating the prediction accuracy of short-term wind energy. In another research work, to improve the forecasting accuracy, a hybrid optimization approach is developed that combines varying mode decomposition (VMD) algorithm, long short-term memory neural-network (LSTM) and firefly algorithm (Qin et al. 2021a, b). Wang et al. (2019) proposed an approach which is the combination of ensemble technique, echo-state-network and wavelet transform, where wavelet transform is utilized to decompose raw wind energy time-series dataset into distinct frequencies with better behaviors and outliers. To automatically learn the input–output non-linear relationship in each frequency and to deal with data noise and model misspecification issues that are common in wind power forecasting problems. Puri and Kumar (2021) used artificial algorithm for wind energy forecasting by utilizing 30 days data of wind speed, air density and temperature in the Himalayan region as input parameters. For multi-class classification, especially with edge classes, Yan et al. (2020) proposed a novel self-weighted robust LDA with 2, 1-norm based pairwise between-class distance criterion, named SWRLDA. Without adding any additional parameters, SWRLDA may automatically skip the optimal mean computation while simultaneously learning adaptive weights for each class pair. Extensive testing shows that SWRLDA outperforms other compared approaches on both synthetic and real-world datasets, while also being more computationally efficient. Chang et al. (2015) developed a new compound rank-k projection (CRP) algorithm for bilinear analysis and applied this algorithm on five different data sets; Coil20, USPS, Pointing'04, CVL and UUIm. The performance of CRP in terms of classification accuracy outperforms that of existing methods, according to experimental results. Zhang et al. (2019) proposed two deep learning-based frameworks for precisely identifying human intentions using innovative spatio-temporal preserving representations of raw EEG streams. Similarly, Luo et al. (2017) evolved a novel semi-supervised feature selection approach. Li et al. (2018a, b) proposed a rank-constrained Spectral clustering (SC) with flexible embedding framework. The authors developed a novel multi-feature adaptive affinity matrix learning technique for SC problem. Li et al. (2019) also offered a novel zero-shot event identification method that takes the advantage of the semantic link between events and ideas. Zhou et al. (2019) proposed a multi-feature fusion with adaptive graph learning model for unsupervised Re-ID (re-identification). It is seen from the literature review that, there are numerous research studies based on the wind power prediction by employing several analytic approaches on various horizons. Previously statistical and persistence methods were utilized. However, for predicting wind turbine power statistical approaches were not first suitable approaches, because these approaches could not handle vast dataset and

could not adapt themselves to nonlinear data and unable forecast long-horizons (Gu et al. 2021; Dong et al. 2021). To achieve this goal, recent studies adopt machine learning algorithms (ML), predominantly, several forecasting models like k-nearest neighbors (kNNs), extra trees, decision trees, deep-learning architectures of LSTM network and logistic regression have been investigated to predict turbine power.

In this paper, the key reason for using machine-learning algorithms is that they can adapt to changes in data patterns and hence create prediction models based on the input dataset rather than using a generic model (Duan et al. 2021; Sun et al. 2019). The key objective of this research work is to forecast short-term wind power based on the past values of wind direction, wind speed, and wind power utilizing algorithms of machine learning and their significance factors on the wind energy parameters are also revealed. In this research, the performance of two ML algorithms such as Gradient Boosting Regression Trees (GBRT) and Support Vector Regression (SVR) are compared to predict the turbine output for the wind farm located in west Turkey. On account of forecasting problems of continuous wind energy values, regression algorithms are implemented. Hence in this research, overall analysis is done in a precise manner to estimate wind energy output. In addition, data analysis framework is also prepared to visualize the data acquired from the Supervisory Control and Data Acquisition (SCADA) system of wind turbines. Dataset analysis has been carried out in Polar and Cartesian coordinates to identify the produced power and wind relationship. Based upon different statistical indices, the performance of the developed model has been evaluated. In addition, the comparative analysis reveals that GBRT model performance is found to be superior than that of the SVR model with Mean-Square Error (MSE) value of 0.026, Root Mean-Square Error (RMSE) 0.301, Mean Absolute Percentage Error (MAPE) 0.063, Mean Absolute Error (MAE) 0.004 and Coefficient of determination (R^2) 0.969. The results obtained illustrate that the proposed model would be applicable in real-life scenarios and help the management team to handle the produced power from wind turbines. The novelty of the research work lies in (a) designing and to construct one-of-a-kind visualization of framework for analyzing data collected by the SCADA system from wind turbines (b) to design and develop an effective machine learning models with Gradient Boosting Regression Trees (GBRT) and Support Vector Regression (SVR) for predicting short-term wind turbine power (c) To quantify the improvements, a comparison study using existing statistical and machine learning methodologies has been performed.

The paper is structured as follows. To forecast turbine power machine learning approach is described in Sect. 2, data pre-processing is discussed in Sect. 3. Section 4 describes applied ML techniques to improve forecast accuracy. Section 5 outlines results and discussions. Finally, conclusion and possible future research directions are given in Sect. 6 of this paper.

2 Machine learning approach to forecast turbine power

2.1 Estimation of input metrological factors

This section presents the estimation of appropriate input metrological factors which will influence the active power generated from wind turbine. The chosen variables i.e. wind speed and wind direction are external input of ML algorithms. The data-set analysis and forecasting is done over the publicly accessible data-set that is acquired from the north-western area of Turkey (Erisen 2018). To collect the data about the power generation and wind properties of the onshore Yalova-region, SCADA system was utilized. The dataset has four input parameters i.e. wind direction, wind speed, active power and theoretical-power. Where, active power is the power produces from wind turbine which is measured in kW, wind speed is measured at the turbine's hub height, measured in m/s, theoretical power is the power value calculated by utilizing speed of wind related to kinetic energy, measured in kWh and wind direction, logs the direction of the turbine that turns automatically to the direction where the wind blows, measured in degrees (°).

2.2 Analysis of SCADA dataset

The Fig. 1 illustrates the steps involved in data analysis. Exploration of data is the primary step in the data analysis where data analyst visually examines a large amount of data to determine the dataset's properties and the points of interest. Data cleaning is to identify the missing part, incorrect, incomplete, inaccurate and irrelevant part of the dataset and then removing, amending and replacing the data according to the requirement. The modelling signifies training the ML algorithm to predict the levels from the assemblies, validating and tuning it for the holdout dataset (Chen and Folly 2018). In addition distinct performance metrics are used to check the performance of ML algorithms by utilizing the training and testing data.

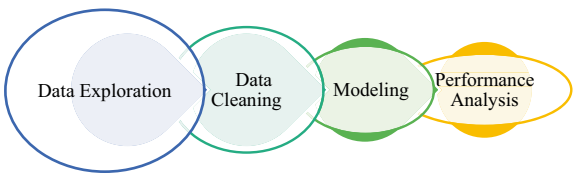


Fig. 1 Predictive analysis steps

For measuring and saving the wind turbine data such as wind direction, wind speed, theoretical power and produced power on the basis of the turbine's power curve, SCADA system is employed. The measurements were made as time series dataset that logs the data at every 10 min for the time period 1st January to 31st December 2018. However the obtained data contain few points where produced output power from turbine is absent, which could be the cause of low speed of wind than cut-in speed of turbine, malfunction (sensor), system processing error and maintenance of wind turbine. The obtained data from SCADA system holds 50,530 observations while 3497 data-points in the dataset are outliers as at these timestamps, power generation is zero. The rest 47,033 observations are utilized for developing the ML models. The dataset is separated into two portions: first 75% of the dataset is considered for training and the rest 25% dataset is considered for testing purpose. Table 1 gives the technical-specifications of the wind turbine and Table 2 presents the statistical indicators of the data-set.

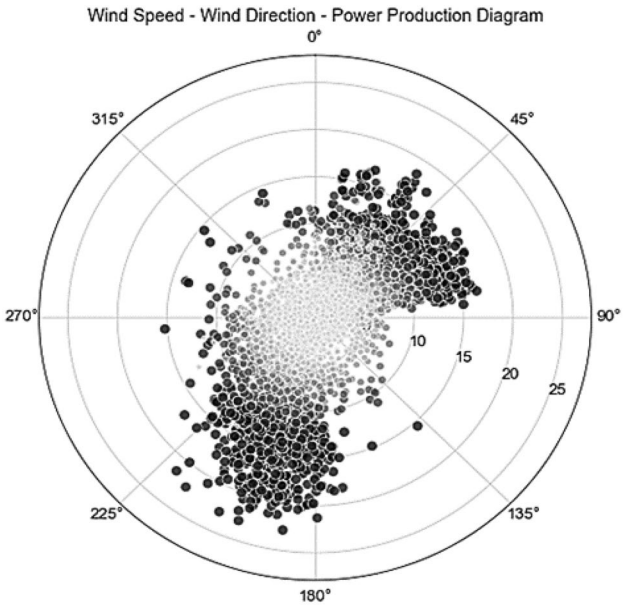


Fig. 2 2D-visualization of wind turbine SCADA data-set in polar coordinate system

2.3 Analysis in polar-coordinates, cartesian-coordinates and pair plot

In the Fig. 2 polar coordinates representing the speed of wind and direction of wind for the whole year. The polar diagram compass rose is utilized as a base model to find the most active regions where the wind blows. Basically, the polar plot shows the qualitative-distribution of produced power with wind speed and wind direction. In the polar diagram it is clearly observable that direction of wind, wind speed and produced power are extensively correlated as wind turbine produces more power when wind blows from the directions between 180°–225° and 0°–90°. In addition, no power is produced above the cut-out speed of 22 m/s. Furthermore, electricity output from certain directions is relatively low. Wind speed and wind direction are important factors to consider when examining the generated power since wind speed affects the quantity of energy to be created and wind direction determines the areas where the wind blows over a certain area. Figure 3 shows a 3-dimensional quantitative depiction of the generated power in the Cartesian-coordinate system for the full year, together with wind direction and speed. From the Cartesian-coordinate

Table 1 Wind turbine technical specifications

Characteristics	Wind turbine
SINOVEL	SL1500/90
Cut-off speed	22 m/s
Rated speed	10 m/s
Cut-in speed	3 m/s
No. of blades	3
Swept area	6362 m^2
Rotor diameter	90 m
Hub height	100 m
Rated power	1.5 MW
Design life	25 years

Table 2 Statistical indicators of the data-set

	Active power (kW)	Wind speed (m/s)	Theoretical power curve (kWh)	Wind direction (°)
Count	50,530	50,530	50,530	50,530
Mean	1307.684332	7.557952	1492.175	123.6876
Std	1312.459242	4.227166	1368.018	93.44374
Min	− 2.471405	0	0	0
25%	50.67789	4.201395	161.3282	49.31544
50%	825.838074	7.104594	1063.776	73.71298
75%	2482.507568	10.30002	2964.972	201.6967
Max	3618.73291	25.20601	3600	359.9976

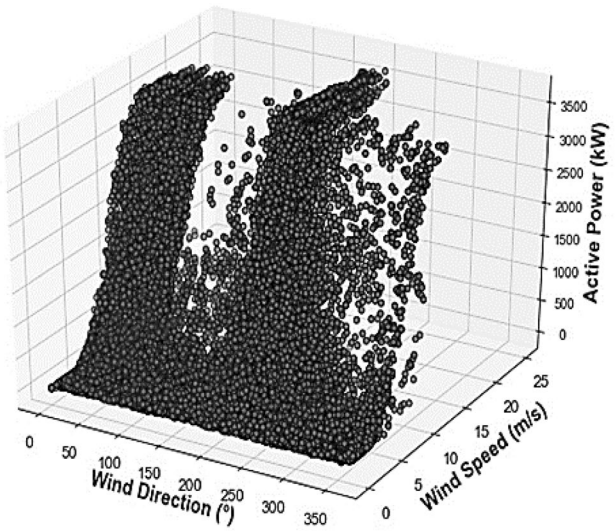


Fig. 3 Scatter plot between wind direction, wind speed and generated power in Cartesian coordinate visualization

plot it is clearly observable that the two regions, where the direction of wind varies from 180° to 230° and 0° to 90° are denser region where contribution of produced power is more than the other regions. Figure 4 shows scatter pair plot that helps to describe the correlation between each feature. The diagonal plots depict histograms which show the probability distribution of every metrological parameters. Whereas, the above and below triangles presents the scatter-plots exhibiting the relationships between the metrological parameters. In addition, every parameter shows the distribution with the other parameter. Basically, the main aim of pair plot is to show the changes in one parameter based on all the other metrological factors.

3 Pre-processing of SCADA data-set

In machine learning, data processing is a crucial step. It entails discarding or modifying data before it is used in order to enhance the system's performance. In the data-set, 10-min sampling rate was selected as to reduce the impact of

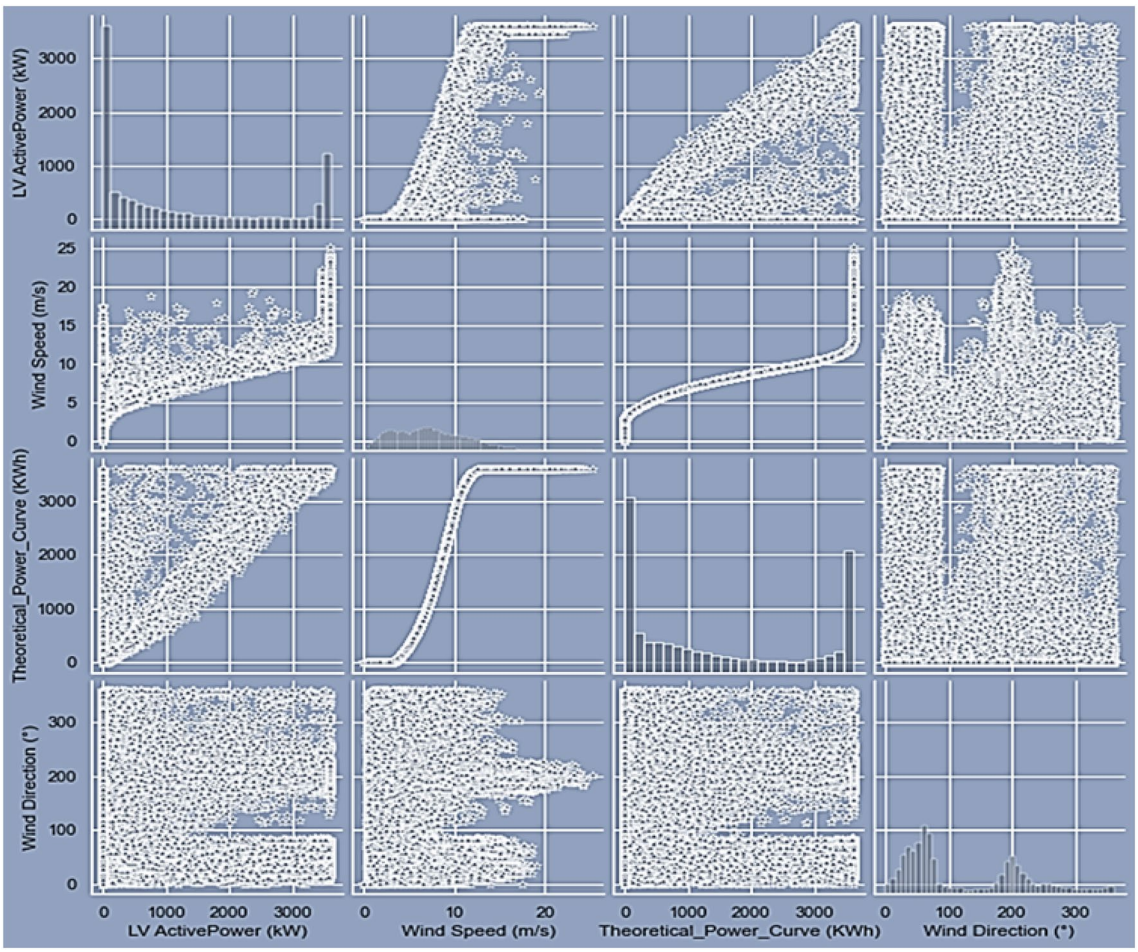


Fig. 4 Pair plot representing the relation between all parameters in the data-set

turbulence and noise. So, in-depth analysis of all input–output features in SCADA dataset, detects some notable errors, for instance, missing values at certain time-stamp, negative values of generated power and wind speed, also zero power generation below the 3 m/s i.e. cut-in speed, has no practical significance in the power production. Hence, those data points including the similar time-stamp are removed from the dataset to limit the negative effect on the prediction models. Such flawed data-points are generally the consequences of maintenance of wind farm, sensor failure and deterioration and error in system processing that makes it important to pre-process the SCADA data-set before the development of prediction models. As the data points in the SCADA system are not of the same order of magnitude so, it is required to bring the data-points in the same order of magnitude. Hence, the given dataset is normalized using the relation.

$$\bar{x} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{1}$$

where, actual data is specified by x,x_{max}(maximumvalueofthedataset), x_{min}(minimumvalueofthedataset) and $\bar{x}$ is the scaled data specified within the range 0–1.

4 Development of machine-learning models and the proposed methods

In this section, implementation of two algorithms i.e. Support Vector Regression (SVR) and Gradient Boosting Regression Trees (GBRT) are implemented. Although, these regression algorithms have dissimilar theoretical background but they provide good forecast performance. Furthermore, these algorithms contains many hyper-parameter that affect the algorithm's forecast performance, run-time, robustness and generalization capability (Abhinav et al. 2017). In addition, trial and error methodology is chosen to identify the best parameters for a particular algorithms, this method is called hyper-parameter tuning.

4.1 Support vector regression

Support vector regression (SVR) was initially proposed by Cortes and Vapnik (1995). It is an extension of support vector machines (SVMs) and is utilized for regression analysis and data classification tasks. This technique has been effectively useful for predicting the wind speed, renewable power production and other modelling tasks. It is used to search the closest regression-plane with all the data points in a set. It gives flexible control on the model complexity and sparse-pattern of solutions, which makes it very effective in solving non-linear problems even with a small sample of training data-sets. SVR also avoids local minima convergence and the problem of over fitting if training period is large (Cortes and Vapnik 1995) In addition, SVR partitioned the dataset into a high dimensional space as follows:

$$\begin{aligned} &\text{Minimize } f(u,\psi) = \frac{1}{2}u^T u + \sigma \sum_{k=1}^{n} \psi_k \\ &\text{Subject to } \begin{cases} \left(Q_k\left(u^T H\left(x_k\right)\right) + h\right) \geq 1 - \psi_k \\ \psi_k \geq 0 \end{cases} \end{aligned} \tag{2}$$

$$\begin{aligned} &F_{Lagrange}(u,h,\psi,\lambda,\gamma) = f(u,\psi) \\ &- \sum_{k=1}^{n} \lambda_k \left(Q_k\left(u^T H\left(x_k\right) + h\right)\right) - 1 + \psi_k - \sum_{k=1}^{n} \gamma_k \psi_k \end{aligned} \tag{3}$$

where γ_k and λ_k are the Lagrange multipliers. Minimisation of (3) by partially differentiating ψ,h and u, respectively, gives

$$u = \sum_{k=1}^{n} \lambda_k Q_k H\left(x_k\right) \tag{4}$$

$$\sum_{k=1}^{n} \lambda_k Q_k = 0 \tag{5}$$

$$\sigma - \lambda_k - \gamma_k = 0 \tag{6}$$

Substituting Eqs. (4) to (6) into (3) forms the dual optimization problem as follows:

$$\begin{aligned} &\text{Maximum } F_{dual}(\lambda) = -\frac{1}{2}\sum_{k,l=1}^{n} Q_k Q_l H\left(x_k\right)^T H\left(x_l\right)\lambda_k\lambda_l + \sum_{k=1}^{n}\lambda_k \\ &\text{Subject to} \begin{cases} \sum_{k=1}^{n} \lambda_k Q_k = 0 \\ \sigma \geq \lambda_k \geq 0, \quad k = 1,2,\ldots,n \end{cases} \end{aligned} \tag{7}$$

The term $H\left(x_k\right)^T H\left(x_l\right)$ in (7) is defined as the kernel function $K\left(x_k, x_l\right)$ and can be expressed in the following way:

$$K\left(x_k, x_l\right) = \exp\left(-\frac{\left\|x_k - x_l\right\|^2}{\varepsilon^2}\right) \tag{8}$$

where ε is the dilation parameter. The parameters, σ and ε in Eqs. (7) and (8), find out the mapping properties of the support vector regression. There was a drop in accuracy on the training data-set for SVR, which began to rise after n = 500. In the algorithm Radial based function was utilized and value of gamma and C were set at 0.01 and 100. Figure 5a presents scatter plot depicting the relationship between power produced (kW) from turbine and wind speed (m/s) and Fig. 5b presents predicted average power as

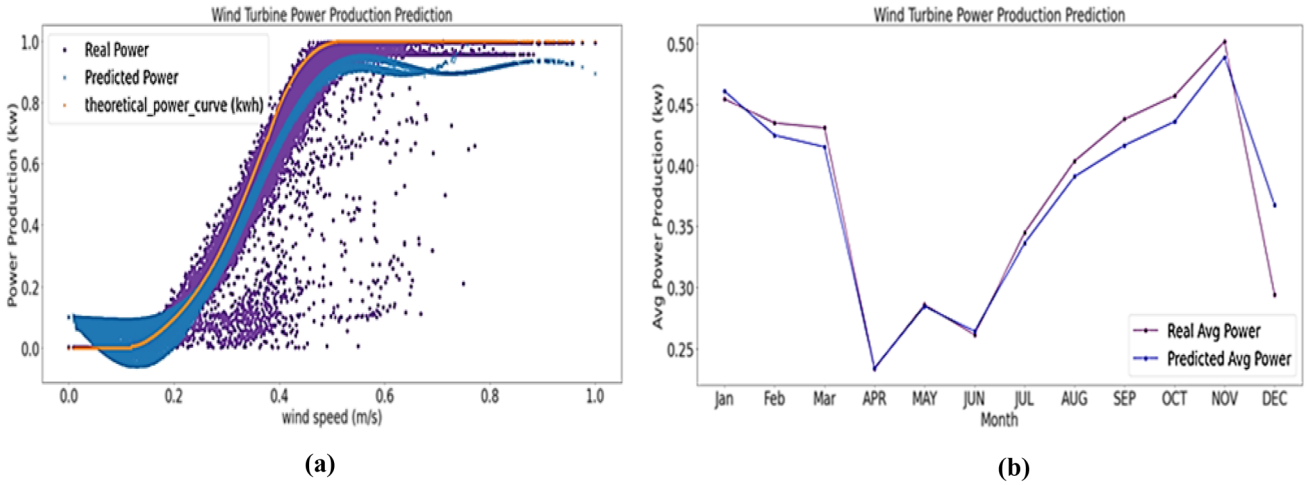


Fig. 5 **a** Scatter plot depicting the relation between power production (kW) from turbine and wind speed (m/s) using support vector regression; **b** Predicted average power compared to real average power from turbine (kW) using support vector regression

compared with the real power from the wind turbine (kW), using SVR algorithm.

4.2 Gradient boosting regression trees

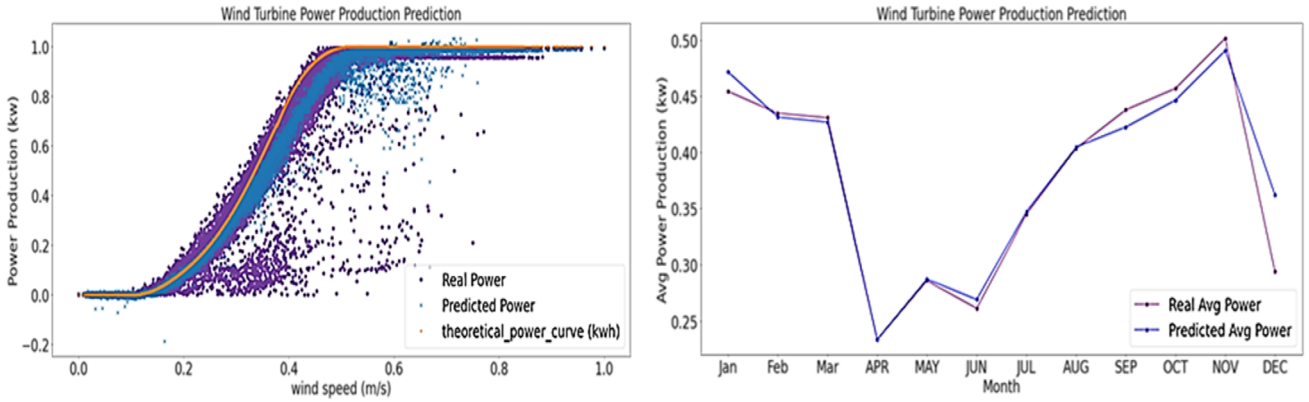
Gradient boosting regression Trees (GBT) is an ensemble-learning algorithm, where vigorous prediction models are designed by the integration of various decision trees or regression trees called as weak-learners. GBT approach minimize error rate with the help of weakly-learned regression models. Weak models have high bias and low variance concerning to the training data and whose responses are examined only approximately improved than random guesses. Basically, boosting algorithms constitute with 3 elements i.e. weak learners, an additive model and a loss-function. GBRT algorithm is the combination of Gradient Boosting trees (GBTs) and Classification and regression trees (CARTs) algorithms. It is observed that performance of CARTs is better in contrast to the Artificial Intelligence (AI) model, as it could model non-linear relationships without necessitating earlier information regarding the probability-distribution of variables. In this algorithm, regression trees are utilised as weak-learners which are produced by CART-algorithm. Weak-learners are introduced to the model to correct the prediction mistakes caused by prior models, lowering the forecasting error and improving model accuracy (Huang et al. 2021). The following is a codified version of the GBRT algorithm.

Algorithm: GBRT.

1. $F_0(x) = argmin_c \sum_{i=1}^{N} L(y_i, c)$
2. For $m = 1$ to M do;
3. $r_{m,i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)}$ $,i = 1,\ldots, N$
4. $c_{m,j} = argmin_c \sum_{x_i \in R_{m,j}} L(y_i, F_{m-1}(x) + c)$
5. $F_m(x) = F_{m-1}(x) + \sum_{j=1}^{J_m} c_{m,j} I(x \in R_{m,j})$
6. $F_M(x) = F_0(x) + \sum_{m=1}^{M} \sum_{j=1}^{J_m} c_{m,j} I(x \in R_{m,j})$
7. $End\ for$

End algorithm

The starting of GBRT algorithm takes place by selecting the initial value of $F_0(x)$, as shown in step 1 of the equation:



(a) **(b)**

Fig. 6 **a** Scatter plot depicting the relation between power production (kW) from turbine and wind speed (m/s) using gradient boosting regression; **b** Predicted average power compared to real average power from turbine (kW) using gradient boosting regression

$$F_0(x) = argmin_c \sum_{i=1}^{N} L(y_i, c) \tag{9}$$

where loss function is $L(.)$ and the following equation is added in step 3 to estimate the value of the negative gradient of the loss-function, as the residual approximation at iteration m, with m = 1,...,M,

$$r_{m,i} = -\left[\frac{\partial L(y_i, F(x_i))}{\partial F(x)}\right]_{F(x)=F_{m-1}(x)}, \quad i = 1, \ldots, N \tag{10}$$

Because the number of splits in each regression tree is assumed to be J_m, so every tree divides the input space into J_m dispoint regions $R_{m,1,\ldots,}R_{m,jm}$ and forecasts a value $c_{m,j}$ for region $R_{m,j}$. The value of $c_{m,j}$ can be estimated through the following equation (step 4):

$$c_{m,j} = argmin_c \sum_{x_i \in R_{m,j}} L(y_i, F_{m-1}(x) + c)$$

The mth regression tree $F_m(x)$, i.e., the updated model, with the leaf node area $R_{m,j},j = 1, 2, \ldots, J_m$ may be derived as follows (step 5);

$$F_m(x) = F_{m-1}(x) + \sum_{j-1}^{J_m} c_{m,j} I(x \in R_{m,j}) \tag{11}$$

where $I = 1$ if $I = 0$ and $x \in R_{m,j}$ otherwise. Furthermore, J_m shows the number of leaf nodes of the mth regression tree. At last, updating of model is done in step 6. This optimization issue is solved using the steepest-descent approach. In the algorithm 100, an_ estimator, maximum tree depth 3 and 0.2 learning rate were chosen. To avoid the problem of overfitting smaller learning rate is selected. Figure 6a shows scatter plot depicting the relationship between power produced (kW) from turbine with wind speed (m/s) using gradient boosting regression trees (GBRT) and Fig. 6b shows the predicted average power as compared with the real average power from turbine (kW) using GBRT algorithm.

5 Results and discussions

On the basis of the comparative analysis carried out in the earlier sections, the key observations and results are scrutinized based on the performance of both the regression models. In addition, selection of various optimal hyperparameters such as regularization of parameters, tree size, and learning-rate are taken empirically via step-wise searching technique for the forecast regression models. The performance of both the forecasting models is estimated using statistical indicators such as Mean-Square Error (MSE), Root Mean-Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Mean Absolute Error (MAE) and Coefficient of determination (R^2). The regression models having less errors indicate the most anticipated and precise algorithm or technique. In addition, among all the performance metrics RMSE is the metric of primary-focus where errors are squared before averaging and put more weightage to large errors. So, low value of RMSE deduces a less error rate in actuality. Table 3 displays statistical indicators for training and testing data-set values for wind power forecasts. Normally, training data-set errors reflect the model's suitability of fitness, whereas testing data-set errors reflect the model's generalization capabilities. To optimize the model performance and accuracy. Parameters of the regression models are tested via hundreds of runs for both algorithms using the max iteration, gamma value, maximum tree depth, C value, number of trees and learning rate etc. Figures 5 and 6 show the performance of the GBRT and SVR regression

Table 3 Data-set statistical indicators

Regression model	Performance evaluation on training-dataset					Performance evaluation on testing-dataset					Training-time (s)
	MSE	RMSE	MAPE	MAE	R^2	MSE	RMSE	MAPE	MAE	R^2	
SVR	0.0357	0.0757	0.0522	0.0483	0.9561	0.0325	0.5745	0.0884	0.04787	0.9471	9.47
GBRT	***0.0260***	***0.0555***	***0.0228***	***0.0031***	***0.9897***	***0.0264***	***0.3012***	***0.0634***	***0.0040***	***0.9690***	***5.83***

Bold and italics indicates better performance

models using overlapping scatter-plots that represent the relationship between wind speed and generated power from the wind turbines as well as a graph comparing predicted average wind power values to actual average wind power generated by the wind turbine, which graphically show the performances of the GBRT and SVR regression models. The outcomes of SVR regression model are displayed in Fig. 5. From Fig. 5a, it is clearly evident that SVR algorithm does not provide good results at high values of wind speed. Also, it is quite observable from Fig. 5b that, this approach could not forecast precise values of wind power as predicted average power curve is not completely overlapping over the real average power curve and takes larger training time as compared to GBRT. Figure 6 presents the results of GBRT regression model. In comparison to the Support Vector regression model, the Gradient Booting regression approach delivers good results in forecasting both low and high values of wind speed and is effective in managing large values of wind speed, as shown in Fig. 6a. Furthermore, from Fig. 6b, it is clearly evident that, predicted average power curve is completely overlapping or well fits over the real average power curve. Therefore, GBRT regression algorithm provides best outcomes as compared to SVR regression algorithm. In Table 3 the performance of SVR and GBRT model is compared on the basis of the performance metrics MSE, RMSE, MAPE, MAE, and R^2 values and it is quite visible from the given performance indices that the GBRT algorithm performs better than the SVR algorithm because it gives less forecast errors. Additionally, GBRT model has higher R^2 values and it takes less training time than SVR Model.

6 Conclusions

Renewable power forecasting is crucial when we deal with the smart-grid and integrating renewable sources into the grid to meet the increasing demand of electricity. Based on the past data power requirements can be forecasted and it is most significant to improve power saving strategies and for managing power production, transmission as well as distribution. In addition, advances in RE power technology have created a slew of new obstacles, and the only way to anticipate the proper power generation is to use new machine learning approaches. Currently, power generated from wind turbine is being used on an immense scale as an alternative source of power. Wind power forecasting is difficult due to the intermittent nature of wind. As a result, integrating wind energy into the main grid is a challenging task. Since wind power forecast can never be accomplished error free, so this provokes the researchers to develop intelligent forecasting models. Based on the data of speed of wind and its direction, a comparative analysis of two regression algorithms is done for wind power prediction. To accomplish the desired aim, the SCADA system data from the Yalova wind-farm which is situated in west Turkey, was collected for the time period 1 January to 31 December 2018 (1 year dataset) at 10-min sampling rate to train and test the SVR and GBRT regression models. The performance evaluation of regression models has been done using various statistical metrics (MSE, RMSE, MAPE, MAE and R^2). The outcomes indicate that SVR and GBRT algorithms are very effective approaches for short-term wind power forecasts. But among the two regression algorithms GBRT exhibits its better accuracy in predicting wind power as compared to that of SVR model with MSE value of 0.026, RMSE 0.301, MAPE 0.063, MAE 0.004 and R^2 0.969. Hence the developed ensemble models are accurate and reliable for hourly forecasts and can be utilized for sustainable balancing and grid integration. The upcoming research areas for further investigation may involve assessment of other deep-learning approaches, advancement of ML algorithms for point-forecasts, predicting combinations and forecast interval-formulation. Future work encompasses the multivariate time-series analysis for wind power forecasting, which will take into account a variety of aspects. In addition, more machine learning methods and meteorological factors should be considered for future studies for more in-depth assessments. In the future, meteorological data could be supplemented with data collected from wind turbines. Because wind turbines can measure and store extra environmental data while generating electricity from wind energy, this knowledge can help with the forecast problem. New features, such as turbine failures, maintenance intervals, and maintenance operations, can be added to the

dataset to expand the amount of features. It may be possible to achieve better forecasting results this way.

References

Abhinav R, Pindoriya NM, Wu J, Long C (2017) Short-term wind power forecasting using wavelet-based neural network. Energy Procedia 142:455–460

Ahmad T, Zhang D, Huang C (2021) Methodological framework for short-and medium-term energy, solar and wind power forecasting with stochastic-based machine learning approach to monetary and energy policy applications. Energy 231:120911

Bo G, Keke L, Hongtao Z, Jinhua Z, Hui H (2021) Short-term forecasting and uncertainty analysis of wind power. J Sol Energy Eng 143(5):054503

Chandran V, Patil CK, Manoharan AM, Ghosh A, Sumithra MG, Karthick A, Rahim R, Arun K (2021) Wind power forecasting based on time series model using deep machine learning algorithms. Mater Today: Proc 47(1):115–126

Chang X, Nie F, Wang S, Yang Y, Zhou X, Zhang C (2015) Compound rank-$ k $ projections for bilinear analysis. IEEE Trans Neural Netw Learn Syst 27(7):1502–1513

Chen Q, Folly KA (2018) Wind power forecasting. IFAC-Papersonline 51(28):414–419

Cortes C, Vapnik V (1995) Support vector networks. Mach Learn 20:273–297

Decaro F, De Stefani J, Vaccaro A, Bontempi G (2021) DAFT-E: feature-based multivariate and multi-step-ahead wind power forecasting. IEEE Trans Sustain Energy 13(2):1199–1209

Delgado I, Fahim M (2021) Wind turbine data analysis and LSTM-based prediction in SCADA system. Energies 14(1):125

Deng X, Shao H, Hu C, Jiang D, Jiang Y (2020) Wind power forecasting methods based on deep learning: a survey. Comput Model Eng Sci 122(1):273

Dong Y, Zhang H, Wang C, Zhou X (2021) A novel hybrid model based on Bernstein polynomial with mixture of Gaussians for wind power forecasting. Appl Energy 286:116545

Duan J, Wang P, Ma W, Tian X, Fang S, Cheng Y, Chang Y, Liu H (2021) Short-term wind power forecasting using the hybrid model of improved variational mode decomposition and Correntropy Long Short-term memory neural network. Energy 214:118980

Erisen B (2018) Wind Turbine Scada Dataset. 2018. Available online: http//www.kaggle.com/berkerisen/wind-turbine-scada-dataset. Accessed 18 May 2020

González-Sopeña JM, Pakrashi V, Ghosh B (2021) An overview of performance evaluation metrics for short-term statistical wind power forecasting. Renew Sustain Energy Rev 138(2021):110515

Gu B, Zhang T, Meng H, Zhang J (2021) Short-term forecasting and uncertainty analysis of wind power based on long short-term memory, cloud model and non-parametric kernel density estimation. Renew Energy 164:687–708

Hanifi S, Liu X, Lin Z, Lotfian S (2020) A critical review of wind power forecasting methods—past, present and future. Energies 13(15):3764

Hu S, Xiang Y, Zhang H, Xie S, Li J, Gu C, Sun W, Liu J (2021) Hybrid forecasting method for wind power integrating spatial correlation and corrected numerical weather prediction. Appl Energy 293:116951

Huang H, Jia R, Shi X, Liang J, Dang J (2021) Feature selection and hyper parameters optimization for short-term wind power forecast. Appl Intell 51:6752–6770

Jacondino WD, da Silva Nascimento AL, Calvetti L, Fisch G, Beneti CAA, da Paz SR (2021) Hourly day-ahead wind power forecasting at two wind farms in northeast Brazil using WRF model. Energy 230:120841

Ju Y, Sun G, Chen Q, Zhang M, Zhu H, Rehman MU (2019) A model combining convolutional neural network and LightGBM algorithm for ultra-short-term wind power forecasting. IEEE Access 7:28309–28318

Kisvari A, Lin Z, Liu X (2021) Wind power forecasting—a data-driven method along with gated recurrent neural network. Renew Energy 163:1895–1909

Li Z, Nie F, Chang X, Nie L, Zhang H, Yang Y (2018a) Rank-constrained spectral clustering with flexible embedding. IEEE Trans Neural Netw Learn Syst 29(12):6073–6082

Li Z, Nie F, Chang X, Yang Y, Zhang C, Sebe N (2018b) Dynamic affinity graph construction for spectral clustering using multiple features. IEEE Trans Neural Netw Learn Syst 29(12):6323–6332

Li Z, Yao L, Chang X, Zhan K, Sun J, Zhang H (2019) Zero-shot event detection via event-adaptive concept relevance mining. Pattern Recogn 88:595–603

Luo M, Chang X, Nie L, Yang Y, Hauptmann AG, Zheng Q (2017) An adaptive semisupervised feature analysis for video semantic recognition. IEEE Trans Cybern 48(2):648–660

Lv J, Zheng X, Pawlak M, Mo W, Miśkowicz M (2021) Very short-term probabilistic wind power prediction using sparse machine learning and nonparametric density estimation algorithms. Renew Energy 177:181–192

Pathak R, Wadhwa A, Khetarpal P, Kumar N (2021) Comparative assessment of regression techniques for wind power forecasting. IETE J Res. https://doi.org/10.1080/03772063.2020.1869591

Puri V, Kumar N (2021) Wind energy forecasting using artificial neural network in himalayan region. Model Earth Syst Environ 8:59–68

Qadir Z, Khan SI, Khalaji E, Munawar HS, Al-Turjman F, Mahmud MP, Kouzani AZ, Le K (2021) Predicting the energy output of hybrid PV–wind renewable energy system using feature selection technique for smart grids. Energy Rep 7:8465–8475

Qin J, Yang J, Chen Y, Ye Q, Li H (2021a) Two-stage short-term wind power forecasting algorithm using different feature-learning models. Fundam Res 1(4):472–481

Qin G, Yan Q, Zhu J, Xu C, Kammen DM (2021b) Day-ahead wind power forecasting based on wind load data using hybrid optimization algorithm. Sustainability 13(3):1164

Sun M, Feng C, Zhang J (2019) Conditional aggregated probabilistic wind power forecasting based on spatio-temporal correlation. Appl Energy 256:113842

Wang H, Lei Z, Liu Y, Peng J, Liu J (2019) Echo state network based ensemble approach for wind power forecasting. Energy Convers Manage 201:112188

Wang H, Xue W, Liu Y, Peng J, Jiang H (2020) Probabilistic wind power forecasting based on spiking neural network. Energy 196:117072

Yan C, Chang X, Luo M, Zheng Q, Zhang X, Li Z, Nie F (2020) Self-weighted robust LDA for multiclass classification with edge classes. ACM Trans Intell Syst Technol (TIST) 12(1):1–19

Yuan R, Wang B, Mao Z, Watada J (2021) Multi-objective wind power scenario forecasting based on PG-GAN. Energy 226:120379

Zhang D, Yao L, Chen K, Wang S, Chang X, Liu Y (2019) Making sense of spatio-temporal preserving representations for EEG-based human intention recognition. IEEE Trans Cybern 50(7):3033–3044

Zhou R, Chang X, Shi L, Shen YD, Yang Y, Nie F (2019) Person reidentification via multi-feature fusion with adaptive graph learning. IEEE Trans Neural Netw Learn Syst 31(5):1592–1601

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

0International Journal of Research in Engineering, Science and Management
Volume 5, Issue 4, April 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Carbon Dioxide Filtering Module

Annoy Roy[1*], Shivani Kumari[2]
[1]*B.Tech. Student, Department of Chemical Engineering, Delhi Technological University, Delhi, India*
[2]*BBA Student, Delhi Technological University, Delhi, India*

***Abstract*: In this paper, we have designed a carbon-dioxide filtering module placed on the rooftops of private and public vehicles in order to purify the ambient air thereby reducing global warming and other pollution related health hazard. In this paper we plan to design and present a solidworks design of our prototype along with few of its applicabilities.**

***Keywords*: Carbon dioxide, filter, pollution, solidworks designing.**

1. Introduction

In this paper, we have designed a carbon-dioxide filtering module placed on the rooftops of private and public vehicles. The module would be adapted enough to filter out pollutants such as PM 2.5 and PM 10 and gases such as CO2 from the ambient air, thus reducing the CO2 levels. The module is designed to filter out a greater quantity of Carbon-dioxide than what is generated by the Vehicle. Once removed from the atmosphere, the carbon would be collected and could be used by ink and plastic manufacturing industries as raw materials.

The air suction inside the module would be carried out in 2 ways- Forced and Automatic. Forced suction would take place when the vehicle is at rest. The module would consist of twin propellers for letting the air into the filter assembly. When the vehicle is in motion, having a certain velocity, automatic suction would take place through the filters, thereby making our product fully energy efficient.

The primary feature of our module includes:

- Energy Efficiency: Our module would rely on its solar panels to power all its electronics
- Carbon Negativity: Since the vehicles equipped with our module would filter more CO2 than they produce, the filtered carbon would be reused.
- Live CO2 Level Monitoring: The CO2 sensors placed within our module would help govt. Monitor live CO2 levels using IoT.
- Easy to Implement: The module does not require any significant technological development and could readily be deployed with the help of major policymakers and companies.
- Low Cost: The module is designed to make it pocket-friendly for the user.

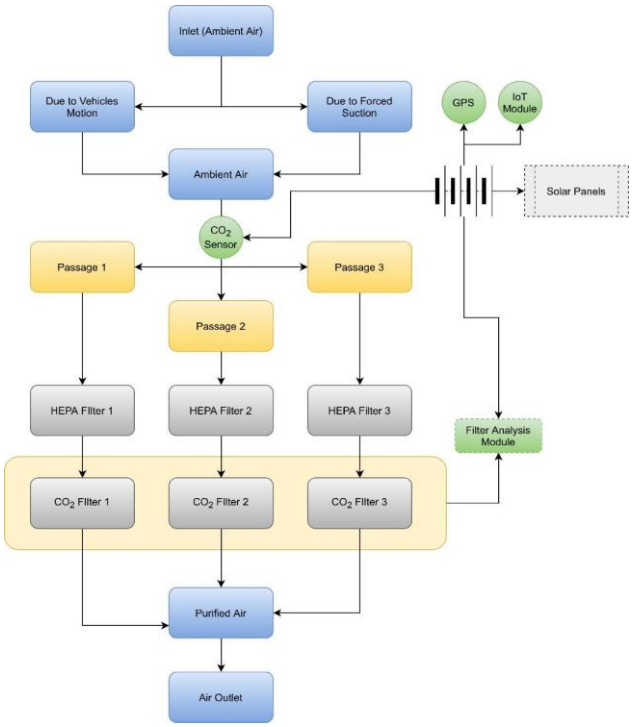


Fig. 1. Block diagram

2. Methodology

The module was designed considering several factors including,

1. *Ease of Integration:* To facilitate the deployment of our module on a large scale it was essential to design it in such a way that it could be used on vehicles of all shapes and sizes. Our module could easily be placed within the roof-top carriers hence eliminating the need of any special fixtures.
2. *Efficiency:* Prime focus was given to make our design highly energy efficient. We have also placed the solar panels on the top surface of the module, together with a battery for storing the clean source of energy. IMU was used to detect the motion of the vehicle, based on the data provided the speed of the suction fans was varied ranging from 0 (when the car vel > 10km/hr) to max (when the car was at rest). This optimized the consumption hence increasing the battery life per cycle
3. *Production Cost:* In order to make the product affordable several modern production methods were considered. We

*Corresponding author: royannoy658@gmail.com

finally decided to use methods like vacuum molding and 3D printing to manufacture the module, as they cut down production time and cost while maintaining the quality.

4. *Serviceability:* Ease of servicing is a major factor when it comes to designing a product like this. We designed the module in such a way that all of its internal components could readily be accessed thereby reducing the service time.
5. *User Comfort:* The user could easily mount and dismount the module from the vehicle. Providing the user, a choice of using the module as and when he seems fit.

3. Modeling and Analysis

We have built a module that will be mounted on the roof of the vehicle and will have standard dimensions, allowing our product to be compatible with the majority of four-wheel vehicles.

It primarily consists of filters, rotors, vibration setup, residue collection section, among other things.

Our module is equipped with 2 modes of suction- forced and automatic. We have added twin propellers for when the vehicle is at rest, they would force in air from the atmosphere to pass in the filter assembly. When the vehicle is in motion, having some certain velocity, the air from the atmosphere would automatically go through the filter, thereby making our module energy efficient and independent.

We have also placed the solar panels on the top surface of the module, together with a battery for storing the clean source of energy, in order to achieve a NET ZERO INITIATIVE, which is a net zero energy initiative.

The rotors will come into action during stationary conditions, and while moving, the rotors slow down there

During the working condition, the carbon residue stuck on the surface of the Carbon filters will be removed with the help of a vibratory setup that uses to and fro motion, and at the end of the process, the sliding collection section located just below the filter will collect the residue with ease, as shown in the diagram. At the end of the process, the waste will be recycled as a raw material for road construction and other uses.

The overall electronics of Breathe is based on the ESP32 module through our custom built PCB. It is connected to the google cloud via Wifi and to our mobile phones via bluetooth. Propeller speed monitor and filter health data is published on the cloud along with vehicle speed and location. This enables us to implement ML models on our custom Firebase server predicting user behavior.

Solar panels provide 50 Watts per hour of energy which is stored in our 7.4V LiPo batteries providing a continuous runtime of more than 10 hrs per charge. The propellers run at a high speed of 20 thousand RPM providing great suction power. The IMU and GPS integrated module provide precise speed and location of our vehicles to be sent to the cloud. Modularity is achieved by using dual CO2 sensors one for Ambient air and the other for Filter health.

1. This is our sensor setup for the initial iteration.
2. The sensor data is published on the cloud using the module.
3. Here is a snip of our code base of sensor and the website.
4. The GUI of our website can be seen here, the graphs and charts provide better readability.

We in our project plan to use a single compound filter with HEPA filter on one side and activated carbon filter on the other. High-efficiency HEPA filter that removes any microscopic pollutant and allergen above 0.3 microns including PM 2.5 and pollen. Activated carbon filter that removes formaldehyde, toxic gasses, VOCs and odor. Activated carbon acts as a carbon dioxide scrubber. Environments with high levels of carbon dioxide benefit from using an air purifier with an activated carbon filter.

We intend to use XAMPP and phpMyAdmin to deliver our co2 sensor data to a MySQL database, from which we will develop a full-fledged web application informing the administrator about the co2 level in a certain area. They can use this data to develop long-term strategies such as increasing public transportation or installing co2 scrubbers in areas with high co2 concentrations.

As a short-term solution, we can connect the traffic signals to a central server using IoT, which can access the sensor's data. When a certain location detects excessive CO2 concentrations, the signals following it might be set red to relieve traffic first, which would enhance air quality.

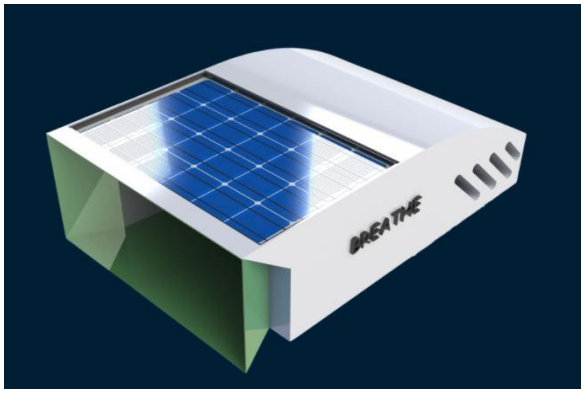


Fig. 2. Breathe

4. Results and Discussion

Whenever the product enters a market, it has several competitors to face and amidst the competitors, the product has to position itself in such a manner that it holds some part of the market share.

Our model can refresh stale air, reducing the chances of health issues caused by outdoor pollutants.

In this project we plan to target:

- Government
- Private companies
- And potential customers.

After evaluating all the necessary steps, we now take a look at the budget. So, the Overall budget of our product is approximately 12,000 which comprises of 3,000 for battery and Solar panels, 3500 for filters, 2500 for 3d printing, 1000 for sensors, 500 for propeller and 400 for microcontroller and 1000 miscellaneous.

5. Conclusion

Creating a vision and mission allows your team to take a top-down approach to your product's development.

Our vision is to reduce carbon dioxide from the atmosphere.

Our mission is to protect people, processes and environment be defining and delivering solutions that combine clean air with energy efficiency in sustainable and profitable way.

References

[1] Christophe Castel, Roda Bounaceur and Eric Favre, “Membrane Processes for Direct Carbon Dioxide Capture from Air: Possibilities and Limitations,” in Frontiers in chemical Engineering, vol. 3, April 2021.

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760992

Comparative Analysis of Time Series Models on COVID-19 Predictions

Puneet Kumar Sehrawat
Department of Information Technology
Delhi Technological University
Delhi, India
puneetsehrawat2000@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological University
Delhi, India
dinesh@dtu.ac.in

***Abstract*— Many Research papers for Covid-19 prediction have been written, where researchers used different models to predict future cases. So, the objective of this paper is to perform a comparative study on all the major models and validate the results obtained before. The analysis will be performed on Indian and American Dataset. The evaluation of all the models will be performed using RMS and r^2error. The forecast models used are ARIMA (Autoregressive integrated moving average), SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors), and Recurrent Neural Network-based LSTM (Long Short-Term Memory) variants like Standard LSTM, Stacked LSTM, Bi-directional LSTM, Convolutional LSTM, GRU (Gated recurrent units) LSTM, and Attention LSTM. These predictive models can offer a crucial insight to policymakers and help normal citizens to prepare accordingly. Among all the mentioned models, GRU LSTM performed the best with a r^2score of 0.986024 followed by Bi-LSTM, Attention LSTM and Stacked LSTM. Furthermore, this research study has also performed the analysis using a multivariate stacked LSTM model which outperformed all the univariate models.**

Keywords*— *COVID-19, Regression, ARIMA, SARIMAX, Long Short-Term Memory Recurrent Neural Network, Time-series Analysis.

I. INTRODUCTION

The current breakout of the coronavirus(COVID or SARS - CoV - 2) which started in the Hubei Province of the People's Republic of China has advanced to the whole world. So WHO Emergency Committee on 30th January 2020, after seeing the spiking increase in the cases in China and all around the world, declared COVID as a global health emergency across the globe[1]. The pathophysiological characteristics of coronavirus have also been defined in a few of the studies that have been conducted, however the mechanism via which COVID spread is highly uncertain[2]. Along with stability in all lands, the present epidemic is affecting almost every aspect of life on a global scale[3][4][5][6]. Therefore, to prevent outbreaks from evolving, more and more research activities are underway to understand the origin, function, treatment and prevention of the COVID virus[7][8][9][10][11]. Various health and government agencies such as the WHO and the American Centers for Disease Control and Prevention (CDC) have provided guidelines and advice on how to prevent the spread of coronavirus changing the way we work and creating a "new normal" for everyone [2].

Government throughout the world proposed multiple lockdowns to prevent the virus from spreading. This virus has high infectivity and is highly contagious, therefore various preventive measures like social distancing, frequent hand washes, wearing masks, and avoiding touching the mouth, nose and eyes. Every effort is being made to tackle this situation and various facilities are also being provided. Hence knowing the number of future cases will prepare us for the future and we can take necessary measures to safeguard ourselves.

In [12] authors applied Gompertz and Logistic, along with Artificial Neural Network on Covid-19 infection to predict inflexion point and several Future cases in Mexico. The R2 for the model (0.9998) is also good but this loss can be improved by using various LSTM Models to remember previous values in an interval. The Authors in [13] used the LSTM (Long Short-Term Memory) model on Canada Dataset and predicted that the possible endpoint of this outbreak will be around June 2020 and there was also a huge dip in the number of new cases around June and July in Canada. Paper [14] showed the effect of weather on the cases of Covid-19, but the effect was not prominent as the number of cases escalated during the period of May-June in India. The Authors in [15] used autoregressive integrated moving average (ARIMA), nonlinear autoregression neural network (NARNN), and LSTM model to forecast the cases in various European Countries and found that the LSTM model performed the best. The method of predicting the future cases is also similar to predicting the stocks prices, so the same techniques can also be used [16], [17].

The objective of this paper is to validate the previously obtained result by performing comparative analysis on Indian and American Dataset. The dataset will be split into 4:1 ratio for training and testing the models. These models will then be evaluated using the RMS and r^2error. The remainder of this paper is organized as follows. Section II lists all the methods which will be used for the analysis. Section III analyzes and discusses the experiment performed. And finally, conclusions are drawn in section IV.

II. METHODS

For performing the comparative analysis, we will be using various regression and LSTM(Long Short Term Memory) models. The Machine Learning models used are Polynomial Regression, ARIMA(Autoregressive integrated moving average) and, SARIMAX(Seasonal Auto-Regressive Integrated Moving Average with exogenous factors), while the

© IEEE 2022. This article is free to access and download, along with rights for full text and data mining, re-use and analysis.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

Deep Learning Models are Standard LSTM, Stacked LSTM, Bidirectional LSTM, Convolutional LSTM, Gated Recurrent Unit(GRU) LSTM, Attention LSTM.

A. Polynomial Regression

Polynomial Regression is the extension of Linear Regression where an independent variable and dependent variable form a curvilinear relationship of n degree Eq 1 shows the equation for polynomial Regression.

$$y = b_0 + b_1 + b_2x^2 + b_3x^3 + \cdots + b_nx^n \quad (1)$$

Where x and y are the coordinates of all the points and b_0, b_1, b_2 are the parameters of the model. In the case of n = 1, this equation becomes that of Linear Regression. It is mostly used to analyze a single output variable and doesn't work when the Independent variable X is dependent.

B. ARIMA(Auto-Regressive Integrated Moving Average)

ARIMA is a statistical analysis model which is used to predict future values based on past values. Over the short-term ARIMA outperform many complex structural models and is most prominent for time series forecasting [15]. In this model, the data are different to achieve stationarity. A model that shows stationarity shows there is constancy to the data over time.

$$y^t = \theta_0 + \varphi_1 y^{t-1} + \varphi_2 y^{t-2} + \cdots + \varphi_p y^{t-p} + \varepsilon^t - \theta_1 \varepsilon^{t-1} - \theta_2 \varepsilon^{t-2} - \cdots - \theta_q \varepsilon^{t-q} \quad (2)$$

In Eq. (2), y^t and ε^t represent the original value and arbitrary error at time step t. $\varphi_a (a = 1, 2, \ldots, p)$ and $\varphi_b (a = 1, 2, \ldots, q)$ are parameters of the model.

C. SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors)

SARIMAX is an updated version of the ARIMA model. The introduction of an external predictor is also called an 'exogenous variable' into the model. To know the value of the variable during the forecasting period the only need is to use an exogenous variable. Just like ARIMA and SARIMA, SARIMAX is also a seasonal equivalent model.

D. Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) network is a type of recurrent neural network which is capable of learning the order of dependence in sequence predicting problems. LSTM units are sometimes also called LSTM memory cells. LSTM architecture solves the problem of vanishing gradient [18]. The input, forget and output gates which are there in LSTM memory block are shown in Fig. 1. The states of the gate are shown in (3)-(8):

$$f_t = \sigma(W_f . [h_{t-1}, x_t] + b_f) \quad (3)$$
$$i_t = \sigma(W_i . [h_{t-1}, x_t] + b_i) \quad (4)$$
$$\widetilde{C_t} = \tan h(W_c [h_{c-1}, x_t] + b_c) \quad (5)$$
$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \quad (6)$$
$$O_t = \sigma(W_o [h_{t-1}, x_t] + b_o) \quad (7)$$
$$h_t = O_t * \tan h(C_t) \quad (8)$$

Where x_t, h_t denotes the input and recurrent information at time t, while C_t denotes the cell state of LSTM. W_i, W_o, W_f represent the weights, b represents the bias. i_t, f_t, O_t represents the input, forget, and output gates and σ is sigmoid activation function [19].

E. Stacked LSTM

Stacked LSTM consists of multiple layers of LSTM with multiple memory cells in those layers which leads to an increase in depth of the neural network. The intermediate layer in an LSTM model gives resulting sequential vectors which are then used as inputs for the next layer in the model. the primary purpose of Stacked LSTM is that for each timestamp it gives an output instead of giving a single output for all of the timestamps in model [20].

F. Bidirectional LSTM

A normal RNN model processes the input in one direction and doesn't take into consideration the value in the future. So, Schuster and Paliwal [21] proposed a bidirectional recurrent neural network that trains in two directions of the sequence i.e., it checks the input from first to the last element and from last to first for every element. It consists of two LSTM layers that are in opposite directions. This allows the network to have both past as well as future information of the sequence at all times.

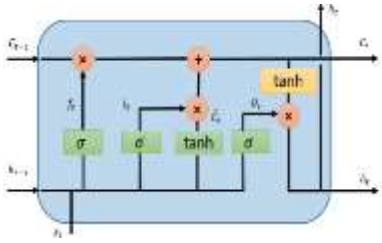


Fig 1. LSTM cell

G. Gated Recurrent Unit GRU

GRU is also a Recurrent Neural Network. GRU consists of an update and a reset gate. The equations in GRU are given in (9)-(12):

Update gate: $z = \sigma(W_z h_{t-1} + U_z h_t)$ (9)
Reset gate: $r = \sigma(W_r h_{t-1} + U_r x_t)$ (10)

Cell state: $c \; or \; \tilde{h}_t = \tan h(W_c(h_{t-1} * r)) + U_c x_t$ (11)

New state: $h_t = (z * c) + ((1 - z) * h_{t-1}$ (12)

Where σ is the sigmoid function. W and U are the parameters of the model while x and h are input and recurrent cell at time t. z represents the output from update gate and c is cell state. The purpose of the reset gate is to define the combination of previous memory with new input and the extent to which the last memory is kept is handled by the update gate [22], [23].

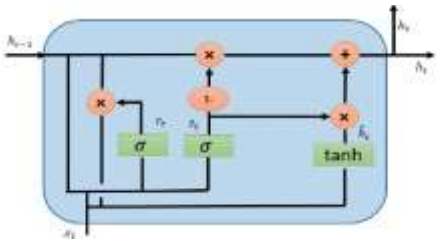


Fig 2. Gated Recurrent Unit (GRU)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

H. Convolutional LSTM (Conv LSTM)

The Convolutional LSTM determines the future state of a certain cell within the grid by using the inputs and beyond states of its local neighbors [19]. Convolutional LSTM consists of vector and gate which are a 3D tension were the last two dimensions act as spatial dimensions. Mathematically It is represented as Eq. (13)-(17).

$$i_t = \sigma(W_{xi} * x_i + W_{hi} * h_{t-1} + W_{ci} \circ c_{t-1} + b_i) \quad (13)$$
$$f_t = \sigma(W_{xf} * x_t + W_{hf} * h_{t-1} + W_{cf} \circ c_{t-1} + b_f) \quad (14)$$
$$c_t = i_t \circ \tan h\,(W_{xc} * x_t + W_{hc} * h_{t-1} + b_c + f_t \circ c_{t-1}) \quad (15)$$
$$o_t = \sigma(W_{xo} * x_t + W_{ho} * h_{t-1} + W_{co} \circ c_t + b_o) \quad (16)$$
$$h_t = o_t \circ \tan h(C_t) \quad (17)$$

where c_t represents the cell state of LSTM. i_t, f_t, O_t represents the input, forget, and output gates, $*$ represents a convolutional operation and $\circ$ represents the Hadamard product.

I. Attention LSTM

Attention LSTM add weights to the previous information and find the indirect correlation between the output and input. The equations are represented in (18) – (20):

$$u_t = tan\ h(w_w h_t + b_w) \quad (18)$$
$$\alpha_t = \frac{exp\,(u_t^T u_w)}{\sum_{k=1}^{T} exp\,(u_t^T u_w)} \quad (19)$$
$$s = \sum_t \alpha_t h_t \quad (20)$$

Where w_w and u_w are weights matrix of attention layer and u_t is the result of the first weighted calculation, b_w represents bias, h_t represents input and α_t is the ultimate weight on h_t to get the final outputs.

III. Experiment and Results

The main aim of this paper is to compare various Machine Learning and Deep Learning Models on the COVID-19 dataset for various countries, like India, United States of America. The dataset used for the study is taken from Kaggle[24], which shows the vaccination progress along with other details about many different countries. Machine Learning Models which were used are Polynomial Regression, ARIMA, SARIMAX and for Deep Learning model will include different variants of LSTM. Performances of all the models are compared using RMSE and r2_score instead of using Accuracy metrics because Covid prediction, just like stocks is an approximation of future curve rather than giving Accurate value. For instance, a curve parallel to actual curve but never intersecting will give zero accuracy but it might still be a better approximation compared to a zig-zag curve touching actual curve at some random points. MinMaxScaler is used to scale the features in a range i.e., between 0 and 1. An 80:20 split is done on the dataset for training and testing purposes as the prediction only works over a small-time.

As you can infer from Tables I, and II that Polynomial Regression, ARIMA, and SARIMAX do not perform very well on the given dataset and therefore generate inconsistent predictions. The r_2 score of polynomial Regression is negative which signifies those predictions are even less accurate than the average value of the dataset over time. ARIMA and SARIMAX also have a negative r_2 score but their RMS error is not that large which shows that ARIMA and SARIMAX perform good in a very short amount of time and then deviate from the correlation. The order of ARIMAX and SARIMAX model was (1, 1, 1) on the Indian dataset with Seasonal Order as (2, 1, 1, 12), and (1, 1, 1, 12) respectively, while for American Dataset

the order was found to be (3, 1, 3) and (3, 1, 1) respectively and Seasonal Order was (1, 1, 1, 12) for both.

As a next step, variants of LSTM are used for predicting the cases. Single Layered LSTM was able to approximate the curve but its RMSE error is greater than other variants. Conv. LSTM more or less gave similar results in all cases when compared with other models. Stacked LSTM, Attention LSTM, and Bidirectional LSTM gave very good results in some cases but average results in other cases. (As you can see in Table I and Table II, Bidirectional LSTM performed good results on the Indian dataset but average in the case of American Dataset). A batch of size 32 is chosen for all the LSTM models along with mean squared error as the loss function and ADAM as the optimizer.

The Standard LSTM used has 50 units of LSTM layer followed by two dense layers of 64 and 32 units. The Stacked LSTM model introduces another LSTM layer between the first LSTM and first dense layer, which contributed to giving better results over standard LSTM. In case of Bidirectional LSTM, the LSTM layer in standard LSTM is replaced by a bidirectional layer with the same 50 units. In convolutional LSTM we have prepared a convolutional model on top of the standard LSTM model. This Convolutional model consists of a Conv1d layer with 128 filters, followed by max pooling, flatten and repeat vector layer followed by the standard LSTM layers. Similarly in case of GRU LSTM, a GRU layer of 50 units is followed by the same standard LSTM layers. The Attention LSTM, just like the stacked LSTM adds another layer just after the LSTM layer, which is Attention Layer in this case.

TABLE I. RMSE AND R_2 SCORE FOR ALL THE MODELS ON INDIAN DATASET

Models	RMS Error	R_2 Score
Polynomial Regression	0.296470	-1.830863
ARIMA (Auto-Regressive Integrated Moving Average)	0.114154	0.580302
SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors)	0.111462	0.599863
Long Short-Term Memory (LSTM)	0.092364	0.839786
Stacked LSTM	0.075742	0.892262
Bidirectional LSTM	0.063920	0.923270
Convolutional LSTM	0.084010	0.867457
GRU(Gated Recurrent Unit) LSTM	0.029104	0.986024
Attention LSTM	0.081056	0.876615

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

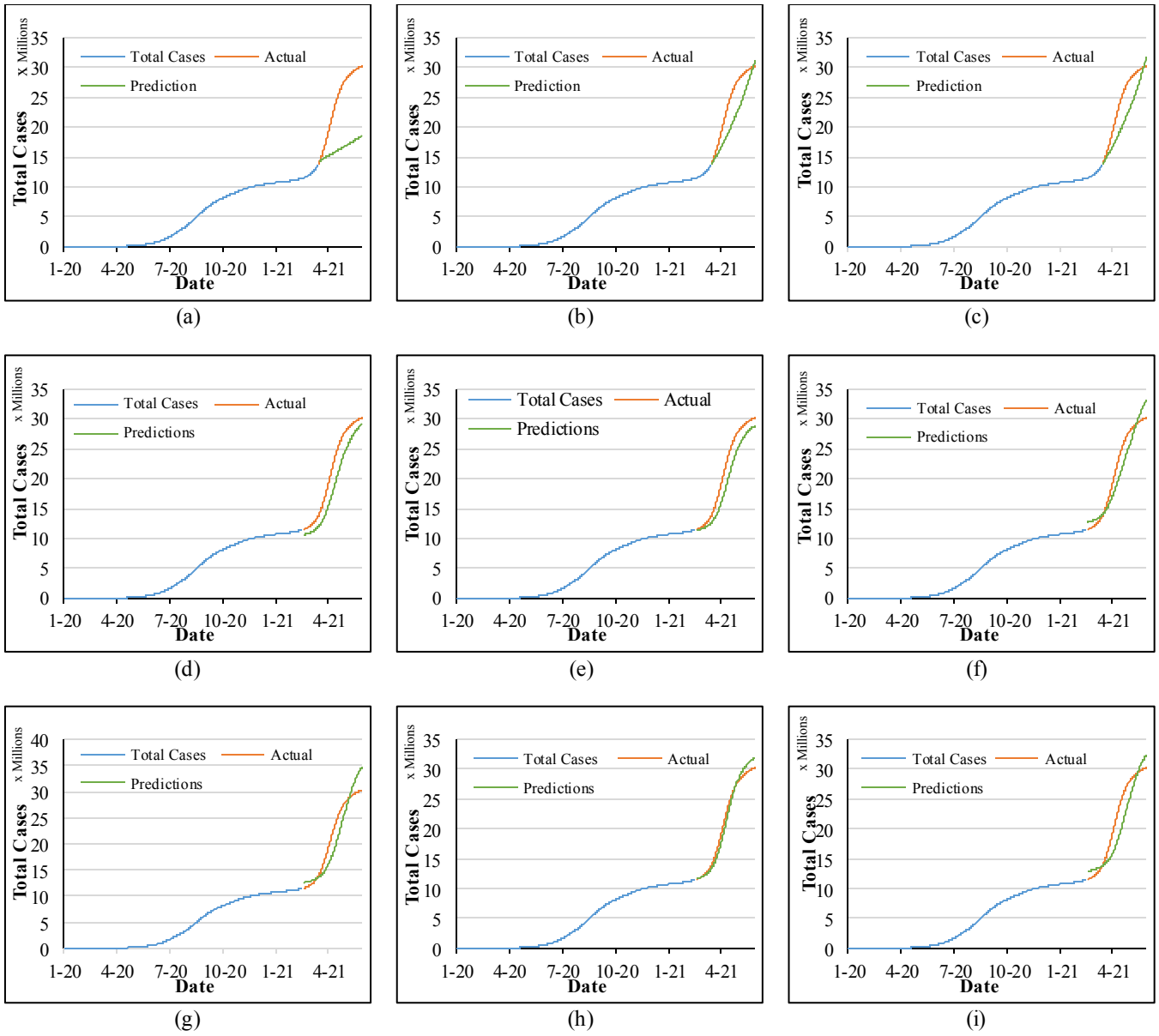


Fig 3. Prediction Plot on Indian Dataset using (a) Polynomial Regression, (b) ARIMA, (c) SARIMAX (d) LSTM, (e) Stacked LSTM, (f) Bidirectional LSTM, (g) Convolutional LSTM, (h) GRU LSTM, and (i) Attention LSTM

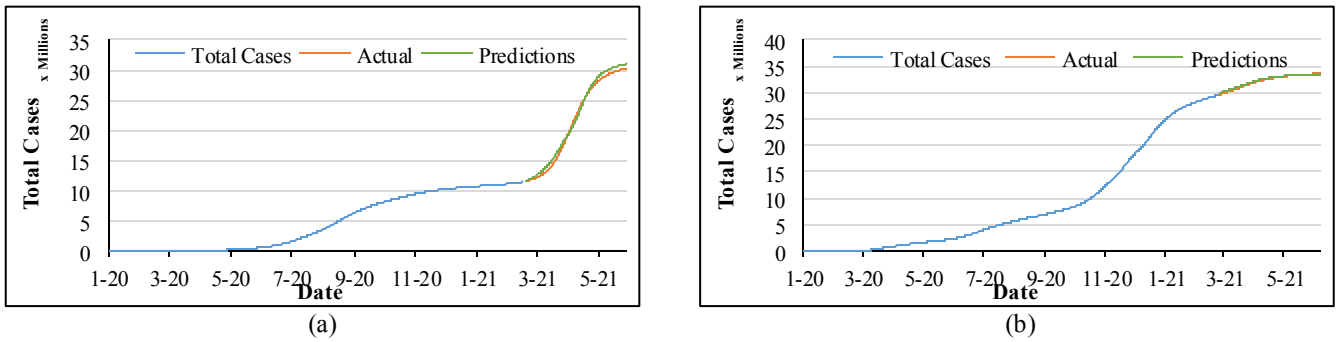


Fig 4. Multivariate LSTM on (a) Indian, and (b) American Dataset

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

Attention LSTM gave a high r_2 score whenever there were no sudden unpredictable future changes in the number of cases due to its ability to focus on segments of sequences. It is able to give its attention to segments of sequences rather than the whole sequence. GRU LSTM gave the best results in almost all the datasets and it was able to predict the cases with very good accuracy. The curve it gave was more or less very close to actual predictions at all times. This is because GRU deals with long-distance information acquisition.

The Plots obtained On Indian Dataset are given in Fig 3 with their respective RMSE and r_2 score given in Table 1. As you can see in the plots GRU LSTM gave a prediction that overlaps with the actual Data. The Curve obtained from Bidirectional LSTM, Conv LSTM and Attention LSTM are the same in shape but due to the variations in Distance from actual cases, The value of RMS error and R Squared Error are different. Bidirectional LSTM can process a sequence of information in both the forward as well as backward direction over a segment of sequence, which is why it is not able to outperform GRU as it needs more information to perform better.

The final Predictions given by Single layered LSTM and Stacked LSTM is better than Bidirectional LSTM, Conv LSTM and Attention LSTM showing that over a long period, Bidirectional LSTM, Conv LSTM and Attention LSTM have some difficulty predicting the cases.

TABLE II. RMSE AND R_2 SCORE FOR ALL THE MODELS ON AMERICAN DATASET

Models	RMS Error	R_2 Score
Polynomial Regression	0.308566	-241.456775
ARIMA (Auto-Regressive Integrated Moving Average)	0.032025	-1.611632
SARIMAX (Seasonal Auto-Regressive Integrated Moving Average with exogenous factors)	0.032882	-1.753269
Long Short-Term Memory (LSTM)	0.019683	0.714537
Stacked LSTM	0.014221	0.850985
Bidirectional LSTM	0.017413	0.776593
Conv LSTM	0.017784	0.766983
GRU LSTM	0.014237	0.928068
Attention LSTM	0.010432	0.919819

Analysis was done using a multivariate model on Indian and American Dataset with parameters as Date, total deaths, total tests, and total vaccinations. This multivariate model due to having various parameters gave the best result on all the datasets. The model used was a simple stacked LSTM model with four layers of LSTM ending with two dense layers of nodes 32 and 1. The optimizer used was ADAM while the activation function for the first Dense layer is RELU. The parameter chosen for this prediction was the number of Deaths, Total Covid Tests, and Number of people vaccinated.

If instead of the Stacked LSTM model some other model like GRU LSTM was used then the accuracy would have been even better as it performed way better than stacked LSTM in univariate data.

TABLE III. RMSE AND R_2 SCORE FOR MULTIVARIATE MODELS ON INDIAN AND AMERICAN DATASET

Model	RMS Error	R_2 Score
LSTM Multivariate on India	0.020758	0.991908
LSTM Multivariate on America	0.005613	0.994082

IV. CONCLUSION

In this paper, analysis of various models was done on the Covid-19 dataset, where Regression models like Polynomial Regression, ARIMA and SARIMAX were not able to get a correlation between the dependent and independent variables. In a short period, SARIMAX and ARIMA also gave good results but the correlation was lost over a long period. RNN based LSTM variants performed better in all cases. Among all the LSTM variants used **GRU LSTM performed the best in Univariate Models** and gave the best results on all the datasets due to its ability to deal with long-distance information acquisition. Attention and Bidirectional LSTM were also able to approximate the cases with good results as long as there were no unexpected rise in cases or any sudden Covid waves.

In the case of **Multivariate data, Stacked LSTM gave R squared error close to 1** showing that as an increase in the number of parameters, the cases also become easier to predict as Active Cases heavily depend on the number of Deaths, Total Covid Tests, and Number of people vaccinated.

So, for future prediction over a long range of time, Univariate models like GRU LSTM can be used as they do not rely on other parameter and still give good results. But for analyzing or predicting the Covid waves or the sudden changes in the curve accurately, multivariate models should be used, for which accurate parameters (like number of Deaths, Total Covid Tests, and Number of people vaccinated) needs to be available, which is not the case over a long period of time. Hence these multivariate models can only be used over a short range of time.

REFERENCES

1. T. P. Velavan and C. G. Meyer, "The COVID-19 epidemic," *Tropical Medicine and International Health*, vol. 25, no. 3. Blackwell Publishing Ltd, pp. 278–280, Mar. 01, 2020. doi: 10.1111/tmi.13383.
2. C. Sohrabi *et al.*, "World Health Organization declares global emergency: A review of the 2019 novel coronavirus (COVID-19)," *International Journal of Surgery*, vol. 76. Elsevier Ltd, pp. 71–76, Apr. 01, 2020. doi: 10.1016/j.ijsu.2020.02.034.
3. K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," in *Children and Youth Services Review*, 2021, vol. 121, no. July 2020, p. 105866. doi: 10.1016/j.childyouth.2020.105866.
4. D. Varshney and D. K. Vishwakarma, "Analysing and Identifying Crucial Evidences for the prediction of False Information proliferated during COVID-19 Outbreak : A Case Study," 2021, pp. 47–51.
5. Yadav and D. K. Vishwakarma, "A Language-independent Network to Analyze the Impact of COVID-19 on the World via Sentiment Analysis," *ACM Transactions on Internet Technology*, vol. 22, no. 1, pp. 1–30, 2022, doi: 10.1145/3475867.
6. K. Saini, D. K. Vishwakarma, and C. Dhiman, "Sentiment Analysis of Twitter Corpus related to COVID-19 induced Lockdown," in *ICSCCC 2021 - International Conference on Secure Cyber Computing and Communications*, 2021, pp. 465–470. doi: 10.1109/ICSCCC51823.2021.9478112.
7. T. Acter, N. Uddin, J. Das, A. Akhter, T. R. Choudhury, and S. Kim, "Evolution of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) as coronavirus disease 2019 (COVID-19) pandemic: A global health emergency," *Science of the Total Environment*, vol. 730. Elsevier B.V., Aug. 15, 2020. doi: 10.1016/j.scitotenv.2020.138996.
8. S. Singh, P. Sapra, A. Garg, and D. K. Vishwakarma, "CNN based Covid-aid: Covid 19 Detection using Chest X-ray," in *Proceedings - 5th International Conference on Computing Methodologies and Communication, ICCMC 2021*, 2021, no. Iccmc, pp. 1791–1797. doi: 10.1109/ICCMC51019.2021.9418407.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

9. K. Chaturvedi, D. K. Vishwakarma, and N. Singh, "COVID-19 and its impact on education, social life and mental health of students: A survey," in *Children and Youth Services Review*, 2021, vol. 121, no. July 2020, p. 105866. doi: 10.1016/j.childyouth.2020.105866.
10. H. Raj and D. K. Vishwakarma, "Detection of Covid-19 in Chest X-ray Image using Convolutional Neural Network," in *2021 2nd Global Conference for Advancement in Technology, GCAT 2021*, 2021, pp. 1–5. doi: 10.1109/GCAT52182.2021.9587725.
11. Kumar, A. Choudhary, A. Kumar, and D. K. Vishwakarma, "Classification and Comparison of Covid-19 Patients," in *Proceedings - 2021 International Conference on Design Innovations for 3Cs Compute Communicate Control, ICDI3C 2021*, 2021, pp. 65–70. doi: 10.1109/ICDI3C53598.2021.00022.
12. O. Torrealba-Rodriguez, R. A. Conde-Gutiérrez, and A. L. Hernández-Javier, "Modeling and prediction of COVID-19 in Mexico applying mathematical and computational models," *Chaos, Solitons and Fractals*, vol. 138, Sep. 2020, doi: 10.1016/j.chaos.2020.109946.
13. V. K. R. Chimmula and L. Zhang, "Time series forecasting of COVID-19 transmission in Canada using LSTM networks," *Chaos, Solitons and Fractals*, vol. 135, Jun. 2020, doi: 10.1016/j.chaos.2020.109864.
14. S. Gupta, G. S. Raghuwanshi, and A. Chanda, "Effect of weather on COVID-19 spread in the US: A prediction model for India in 2020," *Science of the Total Environment*, vol. 728, Aug. 2020, doi: 10.1016/j.scitotenv.2020.138860.
15. Meyler, G. Kenny, and T. Quinn, "FORECASTING IRISH INFLATION USING ARIMA MODELS."
16. H. K. Andi, "An Accurate Bitcoin Price Prediction using logistic regression with LSTM Machine Learning model," *Journal of Soft Computing Paradigm*, vol. 3, no. 3, pp. 205–217, Sep. 2021, doi: 10.36548/jscp.2021.3.006.
17. [C. Anand, "Comparison of Stock Price Prediction Models using Pre-trained Neural Networks," *Journal of Ubiquitous Computing and Communication Technologies*, vol. 3, no. 2, pp. 122–134, Jul. 2021, doi: 10.36548/jucct.2021.2.005.
18. Y. Bengio, P. Simard, and P. Frasconi, "Learning Long-Term Dependencies with Gradient Descent is Difficult."
19. Z. Alom, A. T. Moody, N. Maruyama, B. C. van Essen, and T. M. Taha, "Effective Quantization Approaches for Recurrent Neural Networks."
20. N. Jahromi, S. Hashemi, A. Dehghantanha, R. M. Parizi, and K. K. R. Choo, "An Enhanced Stacked LSTM Method with No Random Initialization for Malware Threat Hunting in Safety and Time-Critical Systems," *IEEE Transactions on Emerging Topics in Computational Intelligence*, vol. 4, no. 5, pp. 630–640, Oct. 2020, doi: 10.1109/TETCI.2019.2910243.
21. M. Schuster and K. K. Paliwal, "Bidirectional Recurrent Neural Networks," 1997.
22. P. T. Yamak, L. Yujian, and P. K. Gadosey, "A comparison between ARIMA, LSTM, and GRU for time series forecasting," in *PervasiveHealth: Pervasive Computing Technologies for Healthcare*, Dec. 2019, pp. 49–55. doi: 10.1145/3377713.3377722.
23. A. Gulli, "Title Page Deep Learning with Keras Implement neural networks with Keras on Theano and TensorFlow."
24. Fedesoriano, "COVID-19 World Vaccination Progress Data," *Kaggle*, Jan. 02, 2021. https://www.kaggle.com/fedesoriano/coronavirus-covid19-vaccinations-data (accessed Jan. 09, 2022).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:02:07 UTC from IEEE Xplore. Restrictions apply.

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Comparative Analysis of Transformer Models on WikiHow Dataset

Dev Jadeja
Department of Information Technology
Delhi Technological University
New Delhi, India
devjadeja559@gmail.com

Avinash Khetri
Department of Information Technology
Delhi Technological Universty
New Delhi, India
avinashkhetri@gmail.com

Anubhav Mittal
Department of Information Technology
Delhi Technological University
New Delhi, India
anubhav11march@gmail.com

Dinesh Kumar Vishwakarma
Department of Information Technology
Delhi Technological Universty
New Delhi, India
dinesh@dtu.ac.in

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9761043

***Abstract*—Because of the ever-increasing amount of information around us, it has become highly essential to consume it in the most efficient way as possible. Text Summarization is a way of generating a summary for a long piece of text while keeping the context and the meaning conveyed intact. Unlike Extractive Text Summarization, Abstractive Text Summarization is the process in which the summary is generated in a more humanlike way, without directly selecting a few sentences from the original piece of text itself. This research study has conducted a comparative analysis of T5 and PEGASUS, which are state-of-the-art abstractive text summarizers. The proposed study has used WikiHow Dataset for the comparison because its summaries are evenly extracted. The evaluation metrics used are BLEU and ROUGE. Unlike other researches doing comparisons, our novelty is in human evaluation for a subset of the summaries to see how reliable the evaluation metrics really are for abstractive summarization and in the use of the WikHow Dataset.**

Keywords—Abstractive Text Summarization, transformer, PEGASUS, WikiHow, T5

I. INTRODUCTION

In recent times the rate at which the amount of new information uploaded to the Internet has risen exponentially [1]. It is becoming burdensome for a normal person to keep up with the events happening around them. Consuming information in it's raw form is not feasible. Thus, it is becoming crucial to extract only the important parts of the information so that it can be consumed in less time [2]. Our main motivation is to review the state-of-the-art summarizers which are based on transformer models on WikiHow Dataset because unlike some other datasets on which the comparison is done, WikiHow dataset has a more evenly extracted out summary from the articles.

Text Summarization is one such way to convert long piece of text into a shorter one without losing the original context and meaning of the content. Mainly there are two ways in which Text Summarization is conducted.

- Extractive Text Summarization – In this technique sentences of the text are given a weightage to judge the criticality and relevance in conveying meaning of the text. The sentences with the most weightage are picked to form the gist [3], [4].
- Abstractive Text Summarization – Unlike Extractive Text Summarization, here the sentences are not directly picked up from the text itself. But a general interpretation of the text is made and the sentences are paraphrased to make the summary more humanlike [5], [6].

There are also many specific use cases for text summarization. Search Engines crawl through the websites to provide a comprehensive title for it to be indexed against. They also extract particular information from webpages that they think the user is looking for [7].



Fig. 1. Use case of Text Summarization by extracting relevant information

Many news media houses use it for headline generation for their articles. Some of the uses also contain lawsuit summarization and medical reports summarization [8].

II. BACKGROUND

The very early approaches of the Text Summarization were extractive in nature. They consisted of methods in which words or whole sentences of higher significance were extracted out of the document to produce a summary because of their inability to understand the context of the data which is provided to be summarized [9].

A. Early Work

Early work in Abstractive Text Summarization can be classified in three parts [10] :-

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:59:02 UTC from IEEE Xplore. Restrictions apply.

- First is the **Semantic-based** approach. It works on finding verb and noun phrases by using semantic illustration present in a text document as an input to natural language formation system. The systems described before includes graph methods, semantic multimodal techniques, text representation in a semantic way and item-based methods [10].
- Second is the **Structure-based** approach which works by encoding the data provided using some logics like templates or structures like graphs, trees, body, rules (lists and classes), lead and body [11].
- Third is the **deep learning-based** approach where sequence to sequence models are the foundational blocks of the studies conducted [12].

B. Transformer Models

Transformer is a novel architecture which is proposed by Google researchers [13]. They are heavily based on attention mechanisms. When experiments were run on tasks which required machine translation, this architecture completed the tasks in lesser time and the results were also better when compared to previous architectures. Two of the transformer-based models which we have used in our research are PEGASUS [14] and T5 [15].

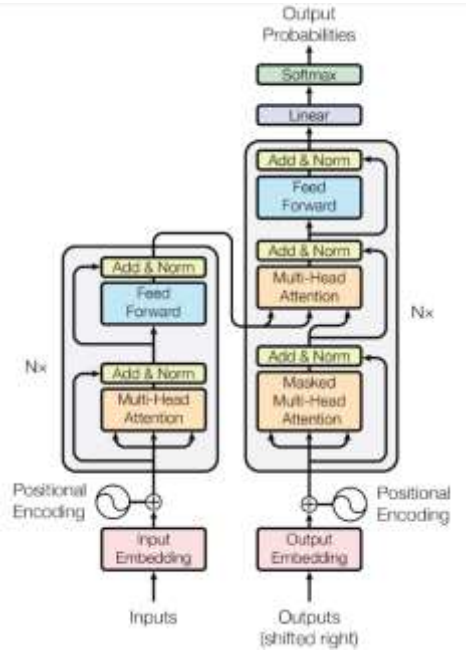


Fig. 2. The Transformer - model architecture [13]

C. PEGASUS

Google researchers recently proposed an abstractive summarization approach called PEGASUS which stands for **P**re-training with **E**xtracted **G**ap-sentences **A**bstractive **S**ummarization **S**equence to sequence models [14]. All the other previous models got out-performed by this model with better results. A goal of pre-training was proposed by them for abstractive summarization tasks. This was called Gap Sentence Generation. They used significant sentence selection as their main strategy. HugeNews and C4(Colossal and Cleaned version of Common Crawl) datasets were used. They used Gigaword [16], CNN/Daily Mail [17], Wikihow, Reddit TIFU and many more for summarization tasks. Evaluation metric used was ROUGE.

D. T5

T5(**T**ransfer Learning with a Unified **T**ext-**t**o-**T**ext **T**ransformer) was also proposed by Google researchers. In NLP they use transfer learning techniques by introducing a framework which converts text-based language problems into a text-to-text format. They used a C4(Colossal Clean Crawled Corpus) as the dataset. They achieved great results on benchmarks like question answering, summarization, text classification and more [15].

III. DATASET

The dataset that we have used is WikiHow [18]. It contains more than 230,000 pairs of articles and summaries. All the content is written by different human authors having different writing styles to avoid bias towards a particular style. All the WikiHow articles are divided in several steps with each step starting with a sentence in bold, collection of whose represents a summary of the article as shown in Fig. 3.

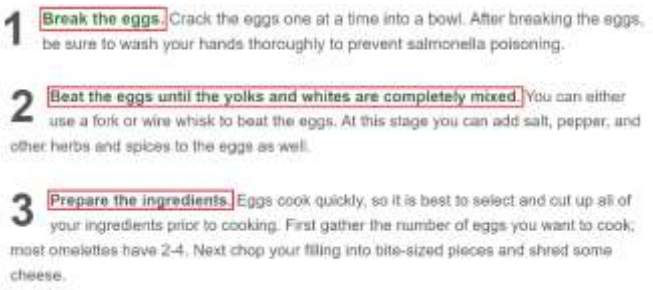


Fig. 3. Depiction of process of creating WikiHow summaries

This is beneficial because the summary is not skewed towards the starting of the article as it is in many other datasets.

IV. METHODOLOGY

This section describes the methodology undertaken so as to achieve the given results. Google Collab was used to carry out the code and we used pre-trained models from Hugging Face. The given Fig. 4 provides a brief insight related to the methodology of the research. The methodology with which we have done the research is quite unique and advantageous because rather than relying on scores from some evaluation metrics, we are also doing human evaluation which acts as a final performance measurer and a reliability check for the evaluation metrics used.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:59:02 UTC from IEEE Xplore. Restrictions apply.



Fig. 4. Flowchart for our research Methodology

A. Data Preparation

We have used data to generate summaries using different models and then tested it on different parameters. We preferred the WikiHow dataset, as mentioned above because the important information is more evenly spread rather than being concentrated in some sections, like in the CNN/Dailymail dataset. Pre-processing included the following :-

- Cleaning - If a summary is greater than .75 times the length of the entire article then it has been removed from the set.
- Lower casing - Convert the input text into the same casing so that all the cases are treated similarly.
- Contraction mapping - Expanding the shorter words such as “don’t” and “isn’t” into words like “do not” and “is not” respectively.
- Removing punctuation - Eliminate punctuation and HTML tags and links, all those things which don't add any meaning to the text such as “!"#$%&'()*+,-./:;<=>?@[\]^_`{|}~”.

The text length distribution is somewhat consistent with the average length of text being 622.742 words. The summary length (of first 1000 entries) has an average length of 47.454.

B. Generating Summaries

- Hugging Face provides us with pre-trained summarization models. We used the google/pegasus-wikihow model and deep-learning-analytics/wikihow-t5-small model. These models are already trained specifically on the WikiHow dataset.
- We took the first 200 articles from the dataset and summarized them using both models. The maximum length for summaries generated by the T5 model was set to 50 because as we saw in the dataset exploration, the average length of the given summaries was 47.454.
- Their results were stored in the dataset CSV file alongside their respective articles.

C. Evaluation and Comparison

ROGUE and BLEU are state-of-the-art metrics for comparison of generated summaries, hence we decided to add them as the evaluation metric. The metrics were imported from Hugging Face. We also added human evaluation as an additional criterion to cross-check the results indicated by ROUGE and BLEU. The summaries generated by both of the models were exported into a CSV and then the file was used to run a script, calculating the ROUGE and BLEU scores

- ROUGE – There are several variants for ROUGE [19], but the basic idea behind them is similar. They assign a single numeric score to a summary that tells us how good it is compared to one or more summaries. The approach that ROUGE takes is comparing n-grams of generated summary to n-grams of the references. N-gram is basically a chunk of n-words. Unigrams are individual words of summaries. The corresponding ROUGE score based on unigrams is called ROUGE-1. It compares the reference summary to the machine produced summary and evaluates the score.
- BLEU – BLEU [20] which stands for The Bilingual Evaluation Understudy Score basically summarizes how close a particularly generated text is to the expected text. A perfect mismatch will have a score of 0 while a perfect match will have a score of 1.0. BLEU score is calculated using a modified n-gram precision and a brevity penalty and those in turn are calculated using reference and candidate lengths. N-gram precision number is the number of words that were common in both the summaries divided by total words in the reference summary. Hence the BLEU score is calculated.

V. RESULTS

A. ROUGE Scores

TABLE I. Various ROUGE Scores of PEGASUS and T5

Model	ROUGE-1	ROUGE-2	ROUGE-L
PEGASUS	0.41	0.28	0.37
T5	0.45	0.32	0.39

B. BLEU Scores

TABLE II. BLEU Scores of PEGASUS and T5

Model	BLEU
PEGASUS	0.33
T5	0.36

C. Generated Summaries

TABLE III. Examples of generated summaries

Title	Pegasus Summary	T5 Summary
How to Be an Organized Artist1	Organize your studio or art space for the sake of efficiency and organization the best way to do this is to organize by what you use every day and by what you use regularly.	Keep your supplies in the same quadrant of your home or studio.Keep your workspace clean and tidy.
How to Create a	Draw the outline of	Decide what you

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:59:02 UTC from IEEE Xplore. Restrictions apply.

Neopoprealist Art Work	your neopoprealist mural on a piece of paper then use a pencil to draw lines on the paper that form the image you wish to paint on the wall.	want your mural to be.Decide on a location for your mural.Choose the right color for your project.
How to Drink Enough Water	Drink plenty of water your body needs water to support proper functioning of organs and cells the kidneys and nervous system are the main targets for toxins in your body this is why its important to drink plenty of water.	Be sure that you have a healthy diet.Ask your doctor about the benefits of eating more than one thing is to do it in order to get fit and be happy.
How to Get Your Friend to Pay More Attention to You	Pay attention to what your friends talk about.Use social media to your advantage.Listen to podcasts and listen to the hosts on the radio.	Use your friends' conversations to figure out what pop culture you want to follow.Watch popular shows on social media and watch a group outing the internet.

D. Human Evaluation

We did human evaluation of 200 summaries and found out that out of all the summaries, 138 summaries generated by the T5 model were better than those generated by PEGASUS model. The other 62 summaries were better when generated by PEGASUS model.

VI. CONCLUSION

After all the results that we observed and after doing the human evaluation on 200 summaries we conclude that T5 model has performed better than the PEGASUS model for the task of abstractive text summarization on WikiHow dataset. Also, BLEU and ROUGE Scores are quite reliable as the human evaluation matched with the results of both the evaluation metrics used.

VII. REFERENCES

[1] K. G. Coffman and A. M. Odlyzko, "Growth of the Internet," in Optical fiber telecommunications IV-B, Elsevier, 2002, pp. 17–56.
[2] L. Abualigah, M. Q. Bashabsheh, H. Alabool, and M. Shehab, "Text summarization: a brief review," Recent Advances in NLP: the case of Arabic language, pp. 1–15, 2020.
[3] R. Ferreira et al., "Assessing sentence scoring techniques for extractive text summarization," Expert systems with applications, vol. 40, no. 14, pp. 5755–5764, 2013.
[4] V. Gupta and G. S. Lehal, "A survey of text summarization extractive techniques," Journal of emerging technologies in web intelligence, vol. 2, no. 3, pp. 258–268, 2010.
[5] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," arXiv preprint arXiv:1602.06023, 2016.
[6] N. Moratanch and S. Chitrakala, "A survey on abstractive text summarization," in 2016 International Conference on Circuit, power and computing technologies (ICCPCT), 2016, pp. 1–7.
[7] M. Allahyari et al., "Text summarization techniques: a brief survey," arXiv preprint arXiv:1707.02268, 2017.
[8] Y. Zhang, "Evaluating the factual correctness for abstractive summarization," CS230 Project, 2019.
[9] A. M. Zaki, M. I. Khalil, and H. M. Abbas, "AMHARIC abstractive text summarization," arXiv preprint arXiv:2003.13721, 2020.
[10] N. Rane and S. Govilkar, "Recent trends in deep learning based abstractive text summarization," Int. J. Recent Technol. Eng., vol. 8, no. 3, 2019.
[11] Sciforce, "Towards Automatic Summarization. Part 2. Abstractive Methods.," Feb. 04, 2019. https://medium.com/sciforce/towards-automatic-summarization-part-2-abstractive-methods-c424386a65ea (accessed Mar. 24, 2022).
[12] O. Klymenko, D. Braun, and F. Matthes, "Automatic Text Summarization: A State-of-the-Art Review.," in ICEIS (1), 2020, pp. 648–655.
[13] A. Vaswani et al., "Attention is all you need," Advances in neural information processing systems, vol. 30, 2017.
[14] J. Zhang, Y. Zhao, M. Saleh, and P. Liu, "Pegasus: Pre-training with extracted gap-sentences for abstractive summarization," in International Conference on Machine Learning, 2020, pp. 11328–11339.
[15] C. Raffel et al., "Exploring the limits of transfer learning with a unified text-to-text transformer," arXiv preprint arXiv:1910.10683, 2019.
[16] A. M. Rush, S. Chopra, and J. Weston, "A neural attention model for abstractive sentence summarization," arXiv preprint arXiv:1509.00685, 2015.
[17] R. Nallapati, B. Zhou, C. Gulcehre, and B. Xiang, "Abstractive text summarization using sequence-to-sequence rnns and beyond," arXiv preprint arXiv:1602.06023, 2016.
[18] M. Koupaee and W. Y. Wang, "Wikihow: A large scale text summarization dataset," arXiv preprint arXiv:1810.09305, 2018.
[19] C.-Y. Lin, "Rouge: A package for automatic evaluation of summaries," in Text summarization branches out, 2004, pp. 74–81.
[20] K. Papineni, S. Roukos, T. Ward, and W.-J. Zhu, "Bleu: a method for automatic evaluation of machine translation," in Proceedings of the 40th annual meeting of the Association for Computational Linguistics, 2002, pp. 311–318.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:59:02 UTC from IEEE Xplore. Restrictions apply.

Applied Intelligence
https://doi.org/10.1007/s10489-022-03416-4

Computer-controlled diabetes disease diagnosis technique based on fuzzy inference structure for insulin-dependent patients

Ankit Sharma[1] · **Nilam**[1] · **Harendra Pal Singh**[2]

Accepted: 18 February 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
The goal of this study is to the development of an automated closed-loop advice system for intense insulin therapy in clinical practice. We use a Mamdani-type fuzzy logic structure to develop an insulin advisory system. The origin of a long-term high glucose level could be due to several biological variables, with delays in insulin manufacturing, absorption, and activity being among the most common. To incorporate these reasons in the glucose-insulin structure, we adopt a non-linear delay system proposed by the author (Nilam and Rathee *Discret Contin Dyn Syst - Ser B* 20(9):3115–3129, 2015). The disturbance in glucose concentration due to increment in the values of delay parameters suggested by authors has been regularized using the controller in the present problems. The four separate experiments were undertaken over the structure that are i) no meal consumption, ii) multiple meal intake in a day, iii) atypical meal input, and iv) uncertainties in the model's parameter. The proposed controller reduces the amplitude of ultradian glucose oscillation as compared to value quoted by other authors (Esna-Ashari et al. *J Med Signals Sens* 7(1):8–20 2017; Soylu and Danişman *Turk J Electr Eng Comput Sci* 26(1):172–183, 2018; Wang et al. *J Biol Dyn* 3(1):22–38, 2009) and yet no control mechanism has been found in the literature for limiting the amplitude of ultradian oscillations of glucose levels. These results support that a fuzzy-based intelligence technique could be used for development of an artificial pancreas of clinical patients.

Keywords Type I diabetes · Artificial Pancreas · Closed loop · Fuzzy logic controller

1 Introduction

Longer persistence of diabetes symptoms or an unbalanced blood sugar level in patients produces the higher the risk of complications. These complications include cardiovascular disease, nerve damage (neuropathy), kidney damage (nephropathy), skin conditions, and more dangerous diseases. Insulin and glucagon are pancreatic endocrine hormones that regulate and maintain blood glucose levels. Insulin is secreted from pancreatic β-cells when the glucose concentration rises too high. These cells promote to absorb enough glucose from the blood to provide the fuel or energy to cells as they require. Insulin also triggers the liver to absorb any glucose that has been left behind. When the glucose level in the blood is too low, glucagon is released by α-cells. It causes liver cells to release stored hepatic glucose into the bloodstream and helps in bringing the glucose level back to its baseline. A person's blood glucose concentration level is consistently higher than the normal range of glycemia, that condition known as Diabetes. Diabetic individuals are majorly classified into two groups: Type 1 diabetes and Type 2 Diabetes. Type 1 diabetes develops when the pancreas cells are destroyed by an autoimmune process, resulting in an utter lack of insulin secretion. Type 2 diabetes develops by a combination of insulin resistance and an insufficient compensatory insulin secretory response. Since the lack of insulin in the blood is a single major cause of raised glucose levels, people with such requirements must regularly manage exogenous insulin with first-hand help or via an automated insulin delivery device.

✉ Nilam
drnilam@dce.ac.in

Ankit Sharma
dtu.ankit.sharma@gmail.com

Harendra Pal Singh
harendramaths@gmail.com

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

[2] Cluster Innovation Centre, University of Delhi, Delhi 110007, India

Published online: 04 May 2022

Insulin administration can thus provide proper glucose management in Type I and Type II diabetic patients. The insulin delivery command for the management of diabetes depends on the open-loop, semi-loop, or closed-loop control techniques. The most common procedure is the open-loop glucose control system, in which doctors prescribe a predetermined dose of insulin subcutaneously without continuous examination of the glucose level. The number of insulin doses varies from patient to patient, and it is injected three to four times a day (traditionally, morning, afternoon, and evening) in patients. However, it is painful because the patient needs to be pricked for several injections throughout the day and less reliable for the patient due to the approximation of the insulin dose. Semi-loop control structures were developed to deliver a precise amount of insulin based on intermittent glucose levels. The patient's varying glucose levels are manually measured at three-hour intervals throughout the day. The critical drawbacks of the semi-closed-loop strategy are the delay in measuring subcutaneous glucose concentration and the difficulty in managing glucose levels throughout the postprandial interval, especially overnight. A closed-loop system is an automated insulin administration device that distributes insulin from an insulin pump based on glucose readings. The system consists of three primary components: a continuous glucose sensing operator, a control technique for computing insulin rate by evaluating blood glucose data, and an insulin secretion device. This method directs insulin delivery according to glucose level, accounting for inherent measurement errors and delays in the body's biological mechanism. However, in some cases, the predicted insulin infusion rate cannot keep the glucose concentration close to the basal value because of the complexity inherent in the mathematical models of diabetes. The highly non linearity, delay in biological phenomena, variability in parameters, and other factors can also contribute to the model's complexity. The delay in insulin action or absorption and secretion are significant causes of unstable glycemia. Non-diabetic people's blood glucose levels fluctuate 10% from their basal glycemia every two hours or so. Insulin must be released during this time, but the exact dosage must be calculated carefully for patients.

Fuzzy logic controller (FLC) [5, 6] is a promising and reliable technique to deal with the uncertainties and disturbances in the human glucose/insulin bio-mechanism. In this approach, the model's inputs are represented in terms of fuzzy numbers with their shape derived from experimental data or expert knowledge. In the last two decades, various kinds of FLC structures have been put forward by experts for the management of diabetes. The variables used as inputs and outputs, the calculation algorithm, and the validation (defuzzification) method differ. The Diabetes Control and Complications Trial (DCCT) findings showed that intensive insulin therapy decreases these diseases' occurrence over the long term. If the patient follows insulin treatment, the physician suggests continuous insulin infusion by subcutaneous route. Grant [7] discussed on FLC approach for insulin therapy and concluded that it provides a simple but dynamic approach to sensitive, safety-critical problems employing monitoring a patient's vital signs. Effective FLC schemes of insulin delivery to diabetic patients are designed by [8, 9] based on a closed-loop strategy. Liu et al. [10] implemented the proposed FLC over the Hovorka-Wilinska (HW) model. Their controller requires two glucose level readings for the updation of insulin bolus delivery values. Mauseth et al. [11] evaluated a fully automated fuzzy logic closed-loop insulin dosing controller that did not require the mathematical form of the diabetic system, and they concluded that it allows clinicians to personalize the dosage. Authors [11, 12] have outlined the interval type-2, fuzzy controller to deal with the disturbance in the augmented diabetic model during the process. Their comparative studies showed that the type-2 fuzzy controller provided a faster response and higher accuracy than type-1 vague under the normal condition and presence of uncertainties in parameters of the model. Gharghory et al. [13] designed a Field Programmable Gate Array of a fuzzy control system that infused the exogenous insulin on the information of current and previous values of glucose concentration. Simulation results show the controller's robustness and ability to determine the required insulin dosage to keep the safe range of glucose levels regardless of meal disturbance. Authors [14–17] developed the advisory system by combining the Proportional- Integral (PI-FLC) and Proportional-Derivative (PD-FLC) controllers to improve the quality and performance of the system. The advantages of the fuzzy- PID controller over the other controllers were that it gives quick response, eliminates errors caused by intense initial circumstances, and restores blood glucose level to its basal amount in an hour. Allam et al. [18] used a fuzzy-based recurrent neural network to determine appropriate insulin dosage. Their control strategy can reduce the postprandial glucose concentration and dynamically respond to different glycaemic challenges. Khooban et al. [19] proposed Swarm optimization tuned Mamdani fuzzy controller over the two-delay mathematical model of glucose-insulin dynamics. The heuristic algorithm was adopted to optimize the input and output membership function parameters and the closed-loop weighting parameters. They concluded that the controller was highly appropriate for feasible applications, like the micro-pump of insulin infusion, due to its suitable efficiency under different robust tests. In this same manner, Abadi et al. [20] designed a fuzzy-PI controller based on a Mamdani-type structure in which for optimizing the membership functions, PI controller, and insulin infusion signal simultaneously. They

adopted a heuristic algorithm, namely particle swarm optimization with linearly decreasing weight. Jain et al. [21] developed a fuzzy logic system based on adaptive cuckoo search optimization algorithm (ACS) for diabetes classification. Experiment results showed that the controller is highly accurate and very sensitive by the ACS algorithm for population size 15 and iteration 50 compared to the FF-BAT algorithm. Reddy et al. [22] introduced FFBAT optimized RBFL classification algorithm by blending FFBAT algorithm with the fuzzy set. The FFBAT algorithm is applied to optimize the rules. These optimized rules are used to classify the diabetes data. They concluded that the controller has outperformed by attaining higher accuracy, specificity, and sensitivity than the existing techniques based on simulation results.

According to research, nonlinear DDEs can better predict glucose-insulin dynamics than mathematical models based on nonlinear ODEs. The significance of the proposed DDE model is that it incorporates delay occurs in insulin secretion, absorption, and endogenous glucose synthesis, which are the primary aetiology of diabetes. The presence of a large amount of blood glucose in the body raises the glucose concentration, while its suppression lowers it down. As a result, the DDE architecture will be used for the controller's design that includes the glucose volume factor. Recently published controller [2, 3, 23] scheme based on a delay model regulates the glucose level within a safe range. However, the oscillation amplitude does not diminish after two hours in the metabolic system. According to the available literature, no control strategy has been developed to maintain glucose levels and reduce the amplitude of ultradian oscillations. Because these oscillations are normal and indicative of healthy insulin dynamics, any effective treatment strategy should strive to keep them in check. The author [1] has devised a delay interval for glucose level management but practically adhering to the proper interval for the delay parameters is tough. We want to look into the influence of disturbance in this interval on glucose concentration and then regulation of glucose level in physiological range with the help of a controller. We used a controller in our model to control glucose concentration even when the delay parameter value was increased, which happens regularly in diabetic patients. In the current problem, a mathematical model has been utilized to model a controller to address the issue of adherence to strict intervals of delay parameters and hypoglycaemia, which are the significant concerns in regulating glucose concentration in instruments employing a closed-loop strategy like artificial pancreas.

The following sections are integral to this document. The first section offers a brief overview of the glucose-insulin regulation system's bio mechanics and highlights the author's contribution to glucose level stabilization using a fuzzy logic controller. The problem statements and control approach are discussed in the second part. The third section introduces the physiological model of the two-delay glucose-insulin control system described by the authors [1]. The model's stability condition, as demonstrated by authors [1], is discussed in the fourth part. In section five, we have presented a new Fuzzy Inference Structure method for insulin delivery in type 1 diabetics. The numerical results are reported in the sixth section. The final section presents the conclusion and mentions the scope of future work.

2 Problem statement and control strategy

Diabetes mellitus type-I is insulin-dependent diabetes mellitus, usually caused by an uncontrolled level of blood glucose. This wild glucose level in patients is typically the result of producing insufficient insulin or losing the ability to absorb insulin by cells. Consequently, an increase of glucose concentration in plasma always is beyond the normoglycaemic average of 90 mg/dl. The problem at hand is developing a better expert control system that maintains the glucose level in 90 to 100 mg/dl. The blood glucose control problem posed is very demanding since the doses given by a physician vary from patient to patient. Furthermore, mathematical models of the insulin-glucose dynamics for a TIDM patient are highly non-linear. These models can be modified by including the factors responsible for the malfunction of glucose-insulin mechanisms like delay in insulin secretion, delay in insulin clearance from blood, and different parameters like diet, exercise, stress, etc. The time delays in insulin release and insulin absorption are common malfunctions in the blood regulatory system of diabetic patients, which have been paid much attention to in the glucose-insulin model. Therefore, we considered the two-delay mathematical model of the glucose-insulin system within the framework of these factors. This mathematical model deals with two-time delays, out of which one occurs in insulin secretion and the second in its absorption and action. The variations in the glucose-insulin dynamics are assessed by using different time delays in insulin's action and absorption. Hence, the advisory/control scheme presented in this work is based on a closed-loop feedback architecture where the controller's purpose was to sustain the glucose concentration at the range of 90-100 mg/dl at various kinds of disturbance scenarios.

The model consists of two major factors: the net rate of glucose concentration and insulin concentration. The net rate of glucose level is balanced by having f_2, f_3, and f_4 functions, and the f_1 function well adjusts the net rate of insulin. Glucose space (v_g) as a single compartment is the most important element in the mechanism of the regulatory system. This variable occurs in all functions other than the

f_4 function. In a healthy adult male of 75 kg (165 lb) with a blood volume of 5 L, a blood glucose level of 5.5 mmol/L (100 mg/dL) amounts to 5 g, equivalent to about a teaspoonful of sugar. The volume of glucose distribution was assumed to be 10 liters in diabetic patients. Hence, the v_g was considered a key factor for glucose control in diabetic patients. The glucose level and its rate are the two input parameters of the controller. Based on the information of these inputs, the controller calculated the value of the output variable that is glucose volume. The controller gives the signal to the artificial pancreas to release a required amount of insulin to maintain glucose level by taking the value of v_g calculated by the controller. To control the sugar level, we selected the v_g parameter of f_3 function to be controlled by the controller, which is described by the following expression:

$$f_3(G) = \frac{G}{C_3(V_g - U)} \tag{1}$$

where U is the control input. G is the glucose level (mg/dl).

3 The two delay-mathematical model for diabetes management

Several papers of delay differential mathematical model of diabetes for diagnosis, treatment, and a deeper description of the deranged system have recently appeared in the literature. The major biological action in glucose-insulin system activity is delayed onset. For example, delay in insulin secretion [24, 25], hepatic glucose production [26, 27] by liver, absorption of glucose by insulin-dependent cells and insulin therapies [4, 28, 29]. Therefore, time delays need to be considered in the modeling of glucose regulatory systems. Recently published review articles [30–33] in glucose-insulin model of delay differential equations have also been referred. In insulin-dependent diabetic patients, inadequate insulin synthesis and delayed insulin release are the primary causes of glucose concentration imbalances. Exogenous insulin infusion could not be used to treat high glucose levels caused by a delay in insulin action and absorption in patients. This can be accomplished in this circumstance by controlling the glucose volume distribution in the plasma compartment. As a result, we postulated a model that could reflect the qualities of insulin functionality and the glucose volume distribution parameter. The formulation of the model of the glucose-insulin regulatory system can be outlined in mathematical equations, which are given as follows [1, 34]:

$$\begin{aligned}\frac{dG}{dt} &= G_{in}(t) - f_2(G(t-\tau_1)) \\ &\quad - f_3(G(t))\, f_4(I(t-\tau_2)) + 150\end{aligned} \tag{2}$$

$$\frac{dI}{dt} = f_1(G(t-\tau_1)) - d \cdot I(t) \tag{3}$$

with the initial conditions $I(0) = I_0 \geq 0$, $G(0) = G_0$, $G(t) = G_0$ for $t \in [-\tau_1, 0]$ and $I(t) = I_0$ for $t \in [-\tau_2, 0]$. Parameter d denotes the self-insulin degradation rate. G_{in} is meal intake rate.

The two essential factors in the diabetic model are glucose and insulin, denoted by G(t) and I(t) as glucose and insulin concentration at time $t \geq 0$, respectively. In the body, glucose concentration increases for two main reasons: food intake and glucose production by the liver, known as hepatic glucose production (HGP). The term G_{in} denotes the patients' glucose intake rate. When the glucose concentration starts rising, the β- cell located in the pancreas starts releasing insulin. Insulin helps glucose to enter the fatty tissues and muscles. As a result, the glucose level starts decreasing. When the plasma glucose concentration level drops, α-cells, located in the pancreas, start releasing another hormone called glucagon. Glucagon exerts control over vital metabolic pathways in the liver and sends the signal to the liver for glucose dispense. The model includes two significant time delays: first delay in insulin secretion by the pancreas that is denoted by τ_1 and second delay in bio-mechanism; denoted by τ_2, which means the delay in insulin absorption and insulin action in the body to decrease the increased glucose level. In the model (2)-(3), the biological meaning of the functions $f_1.f_2$, f_3 and f_4 are given as:

$f_1(G)$ stands for insulin production triggered by glucose levels. The most important element for insulin secretagogue is glucose. The usual biphasic pattern of insulin production is induced by rapidly increasing glucose levels. This pattern of secretion disappears after the variations in glucose levels become stable. This pattern appears to be sigmoid in shape.

$f_2(G)$ describes glucose utilization by insulin-independent cells and tissues. There are brain and nerves having the ability to absorb glucose without using insulin. Except in the case of hypoglycaemia, this absorption is not affected by blood glucose levels.

$f_3(G)$ denotes the relationship between insulin-dependent glucose utilization and glucose concentration over the normal physiological range of insulin. The insulin-dependent glucose utilization is mostly due to muscles, fat cells, and other tissues. This function passes through origin when glucose absorption by the brain is subtracted from net glucose utilization and indicates that insulin-dependent glucose uptake is proportional to glucose level.

$f_4(I)$ shows the dependence of glucose utilization on insulin concentration at the basal range of glucose concentration. In general, minor changes in plasma insulin concentration or insulin receptor function causes appreciable alterations of glucose level in the glucose-insulin

metabolism system. The utilization of glucose by insulin-dependent cells follows sigmoid dynamics on the presence of insulin.

The representation of functions are given below in terms of parameters:

$$f_1(G) = \frac{Rm}{1 + exp((C_1 - G/V_g)/a_1)} \tag{4}$$

$$f_2(G) = U_b(1 - exp(-G/(C_2 V_g))) \tag{5}$$

$$f_3(G) = \frac{G}{C_3 V_g} \tag{6}$$

$$f_4(I) = U_0 + \frac{(U_m - U_0)}{1 + exp(-\beta log(I/C_4(1/V_i + 1/Et_i)))} \tag{7}$$

In given functions, G and I are the variables of function and rest of them is constant. The biological explanation of parameters and variables are given as:

R_m: Maximum insulin production rate by pancreas.

C_1: Concentration of glucose at which the pancreas is most effective

V_g: Volume of glucose space (plasma and the intracellular space).

a_1: The scaling factor determines how sensitive the pancreas is to variations in glucose.

U_b: Brain and nerve cells' maximum glucose consumption rate.

C_2: Scaling factor; controls the sensitivity of central nervous system cells to variations in glucose.

C_3: Scaling factor; controls the sensitivity of muscle cells to glucose fluctuations.

U_m: Maximum-insulin consumption rate for glucose utilization by muscles.

U_0: Minimum-insulin consumption rate for glucose utilization by muscles.

β: Parameter is used for employing the function f_4 shape.

C_4: Scaling factor; controls the sensitivity of muscle cells to insulin fluctuations.

V_i: Volume of insulin distribution in interstitial fluid compartment.

E: Insulin diffusion rate.

t_i: Time for self insulin degradation in interstitial fluid compartment.

We assumed the following conditions in the model (2)-(3).

(H_1) $G_{in}(t) \in C(0, \infty)$ are positive functions for t ≥ 0.

(H_2) $f_1(x)$, $f_2(x)$, $f_3(x)$, $f_4(x) \in C^1[0, \infty)$ are positive for $x > 0$. $f_1(0) > 0$, $f_2(0) = f_3(0) = 0$, $f_4(0) > 0$

(H_3) $f_1'(x)$, $f_2'(x)$, $f_3'(x)$, and $f_4'(x)$ are positive on $(0, \infty)$.

Figure 1 shows the shape of functions f_i, i = 1,2,3,4:

4 Analysis of model's stability

The model's unique steady state is specified as (G^*, I^*) and is obtained by solving the $\frac{dG}{dt} = 0$ and $\frac{dI}{dt} = 0$ equations. However, the straightforward solution of G and I of the model (2-3) is quite difficult to find. So, to examine the non-linear model's stability at equilibrium locations (G, I), we choose the linearized approach method [35]. The linearized system of the model (1-2) about the steady point (G^*, I^*) is given by

$$\frac{dG}{dt} = -PG(t) - QI(t - \tau_2) \tag{8}$$

$$\frac{dI}{dt} = R(G(t - \tau_1)) - d \cdot I(t) \tag{9}$$

where $P = f_2'(G^*) + f_3'(G^*) f_4(I^*) > 0$, $Q = f_3(G^*)$ $f_4'(I(t)) > 0$, $R = f_1'(G^*) > 0$

The characteristics equation is given as

$$\lambda^2 + (P + d)\lambda + Pd + QRe^{-\lambda(\tau_1 + \tau_2)} = 0 \tag{10}$$

The positivity and boundedness of the glucose-insulin feedback model's solution have been discussed by the authors [1]. They had also stated results for the existence of a positive solution of the model and its stability and following conclusions have been drawn:- (R_1) If $\tau_1 = 0$ and $\tau_2 = 0$, then (G^*, I^*) is stable.

(R_2) If $\tau_1 > 0$ and $\tau_2 = 0$, and

(2a) If $Pd > QR$, then (G^*, I^*) is stable.

(2b) If $Pd < QR$, then $\exists\ \tau_{1_0} > 0$ such that (G^*, I^*) is stable when $\tau_1 \in (0, \tau_{1_0})$ and unstable when $\tau_1 \geq \tau_{1,0}$.

(R_3) If $\tau_1 = 0$ and $\tau_2 > 0$, and

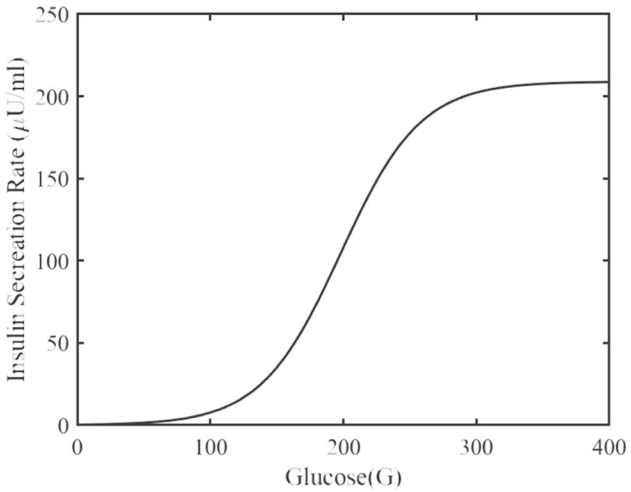
(3a) If $Pd > QR$, then (G^*, I^*) is stable.

(3b) If $Pd < QR$, then $\exists\ \tau_{2_0} > 0$ such that (G^*, I^*) is stable when $\tau_2 \in (0, \tau_{2_0})$ and unstable when $\tau_2 \geq \tau_{2,0}$. (R_4) If $\tau_1 > 0$ and $\tau_2 > 0$, the steady state (G^*, I^*) of the linearized system (8) and (9) is stable for $Pd > QR$.

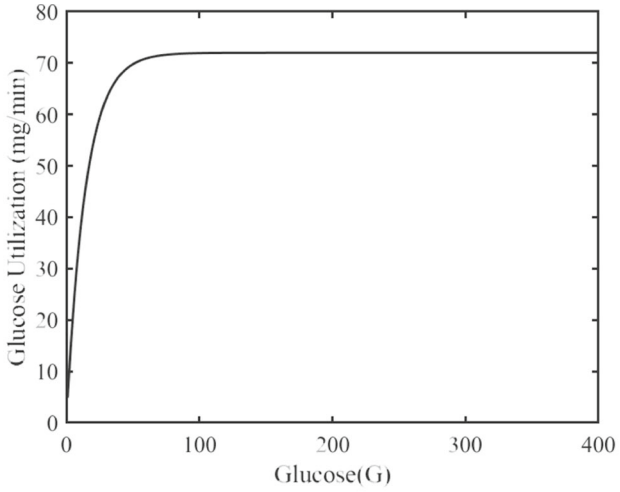
5 Controller's methodology

The Fuzzy optimal controller based on Mamdani fuzzy system is structured with two input linguistic variables and one output linguistic variable. The glucose level G(t) and its rate of change $\frac{dG}{dt}$ are taken input linguistic variables. The output linguistic variable is represented as the insulin volume in FLC. Tables 1 and 2 show the characteristics of input and output variables, respectively.

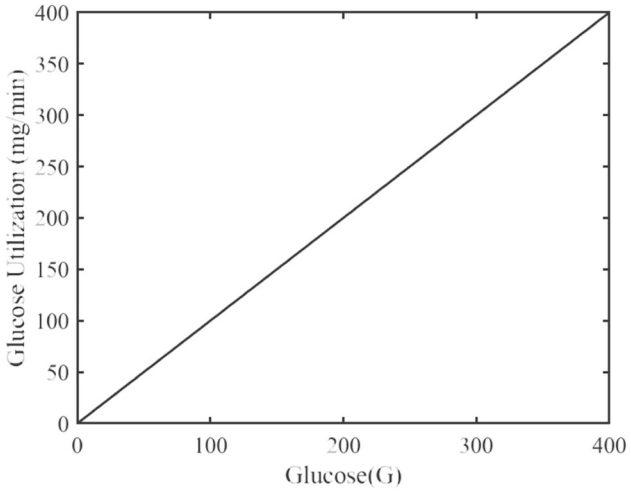
The graphical representation of these variables is shown in Figs. 2 and 3, respectively. The membership functions of



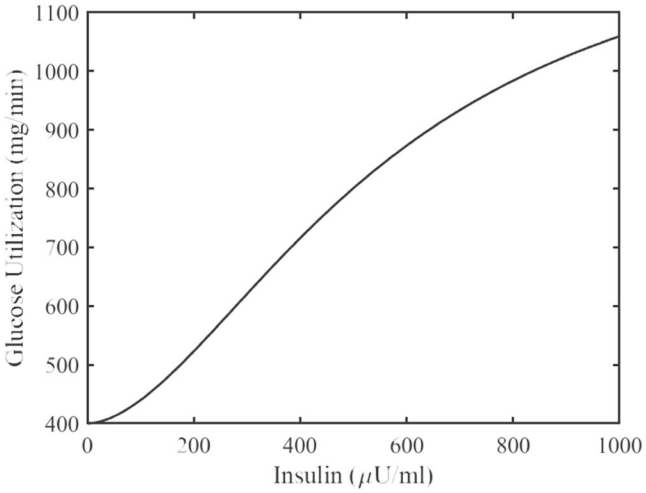
(a). $f_1(G)$: Glucose effect on insulin secretion



(b). $f_2(G)$: Glucose utilization by insulin-independent cells and tissues



(c). $f_3(G)$: Insulin dependent glucose utilization when plasma insulin is 4.5 $\mu U/ml$



(d). $f_4(I)$: Insulin dependent glucose utilization when plasma glucose is 90 mg/dl

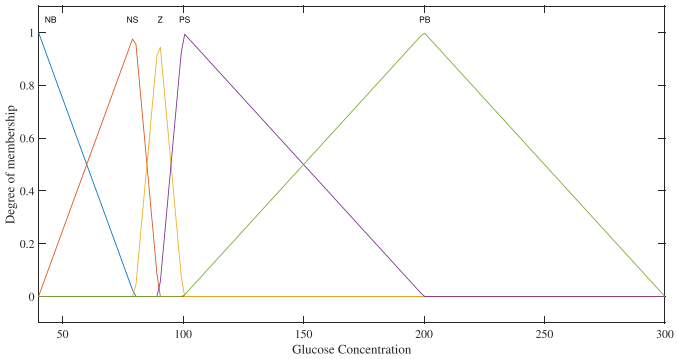
Fig. 1 Graphical representation of f_1, f_2, f_3 and f_4 functions

Table 1 Characteristics of input variables

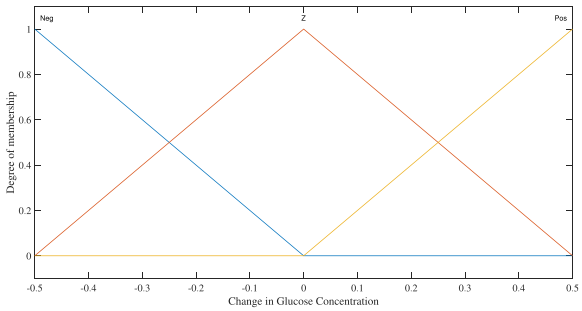
Input variables	Interval	Membership functions				
Glucose Level G(t) (mg/dl)	[40 300]	Negative Big	Negative Small	zero	Positive Small	Positive Big
Glucose Deviation dG/dt	[-0.5 0.5]	Negative	Zero	Positive		

Table 2 Characteristics of output variables

Output Variables	Interval	Membership Functions				
Volume of glucose (v_g)	[-15 5]	Negative Big	Negative Small	Zero	Positive Small	Positive Big



(a). Glucose Concentration



(b). The rate of Change in Glucose Con.

Fig. 2 Membership Function of Input Variables

fuzzy sets are taken according to the fuzzy classification of the input and output variables. The triangular membership function is chosen to define the FLC.

The total number of 15 IF-THEN rules were defined based on the structure of input and output fuzzy sets. These rules were of AND (minimum) type antecedent to minimize the error between the desired and actual glucose concentration. The IF-THEN fuzzy rules were linked to input and output MFs. The CENTROID defuzzification method was taken to find a crisp value. Table 3 shows the fuzzy rules of the controller.

The Fig. 4 shows the surface view of control action.

6 Numerical simulations

Numerical simulation of the two-delay model was carried out using DDE23 solver in MATLAB 2012b. Insulin secretion in the glucose-insulin dynamics occurs in an oscillatory manner over a range of (50 - 150) min in a normal person. Authors [1] observed the same ultradian oscillation of insulin secretion over a range of (50 - 200) min in type 1 diabetic people. In the range of (50 - 200) min, the glucose oscillates from 105 mg/dl to 145 mg/dl for type-I diabetic people. The ultradian oscillations of insulin secretion are assumed to result from an instability in the glucose-insulin endocrine dynamical system. After that, they drew our attention to the range of τ_1 and τ_2 for which normal glucose concentration range arose in people with type 1 diabetes using an artificial pancreas. For τ_1 from 0 to 15 min, and τ_2 was 0, the glucose concentration is slightly out of the reasonable physiological range. The other results of their studies are that the glucose concentration crossed the normal range for a time lag of more than 43 min in τ_2(when τ_1 = 5 min) and a delay of more than 38 min in τ_2 (when τ_1 = 10 min) which may be causing the condition of either hyper or hypoglycaemia, which may lead to serious consequences like kidney failure, diabetic coma, etc., in patients. Author [1] has established an interval for delay for maintenance of glucose level. It is tough to follow the strict interval for the delay parameters. We want to study the effect of distribution in this interval on glucose concentration. We have applied a controller in our model is control glucose concentration even in case of the increased value of delay parameter, which may frequently occur in practice. Since we have the time range of both the delays in which the glucose concentrations become out of safe range, we have implemented the proposed FLC controller over the model where values of delays have been considered beyond its typical values. The following two cases are considered in diabetic patients and tested for an advisory control system.

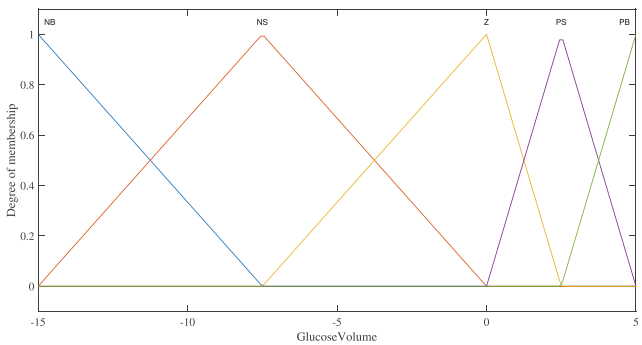


Fig. 3 Membership function of output variable (Glucose Volume)

Table 3 Fuzzy rules

	Change in glucose rate		
Glucose level rate	Negative	Zero	Positive
NB	NB	NB	NB
NS	NS	NB	Z
Z	Z	Z	PS
PS	PB	PB	PS
PB	PS	PS	PB

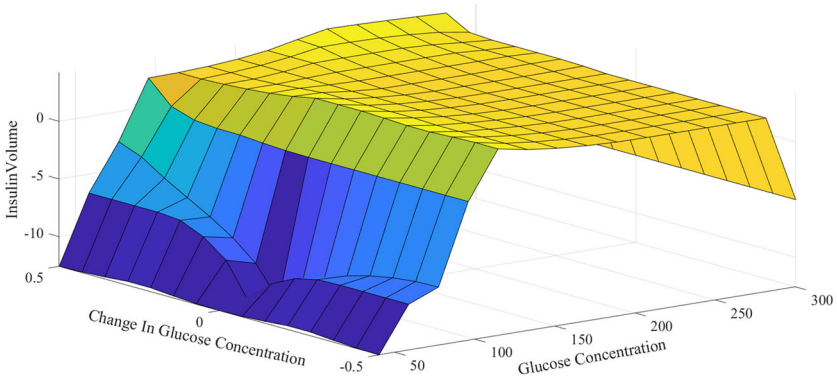


Fig. 4 Membership function of output variable (Glucose Volume)

Case 1. Delay in insulin secretion in diabetic patients ($\tau_1 \geq 0$ and $\tau_2 = 0$)

In this scenario, delay happens in insulin secretion in diabetic patients, but there is no delay in insulin absorption by cells and tissues. The delay τ_1 represents the time lag from food intake to secretion of insulin. The authors [1] concluded that glucose level comes to normal range when $\tau_1 \in (5, 15)$ min. With the τ_1 coming out of this period, glucose concentration goes out of the normal range. That's why in this paper, in addition to already calculated values of time delay, we have considered more than 15 min delay in insulin secretion for simulation. Therefore, the present study's assumed time delay set of τ_1 is 20,25,30 and 35 min.

Case 2. Presence of delay in both insulin secretion and insulin action ($\tau_1 \geq 0$ and $\tau_2 \geq 0$)

Authors [1] considered two cases: In first case, τ_1 was 5 min and τ_2 was within period (31, 43) min. In the second case, τ_1 was equal to 10 min, and τ_2 was within the range of (26, 38) min. They have simulated the two-delay glucose-insulin model. They conclude that the ultradian oscillations range of glucose concentration for each period 182 min after ignoring the first oscillation in both cases were safe for diabetic patients. The summarized results of the author are that the normal range of glucose level is achieved in diabetic patients when $\tau_1 = 5$ min and τ_2 is from 31 to 43 min, while τ_1 = 10 min, the range of τ_2 is from 26 to 38 min. For simulation, the values of the parameters given by author [25] were adopted, which are illustrated in Table 4 for a diabetic patient.

We tested the controller's performances under the four different setups, including without meal disturbance, with meal disturbance, unusual meal intake, and uncertainty in the insulin clearance rate. The explanation of all four setups is as follows:

- Without meal disturbance: In this case, it assumed that the patient has not consumed any food for 800 min.
- With meal disturbance: In this case, it assumed that the patient consumed the meal at a different rate within a specific period.
- Unusual meal intake: In this case, it assumed that the patient consumed the meal within 5 min at any time.
- Uncertainty in insulin clearance rate: In this case, it assumed that insulin clearance rate differs in patients. So, the ±20% uncertainty in this parameter has been assumed.

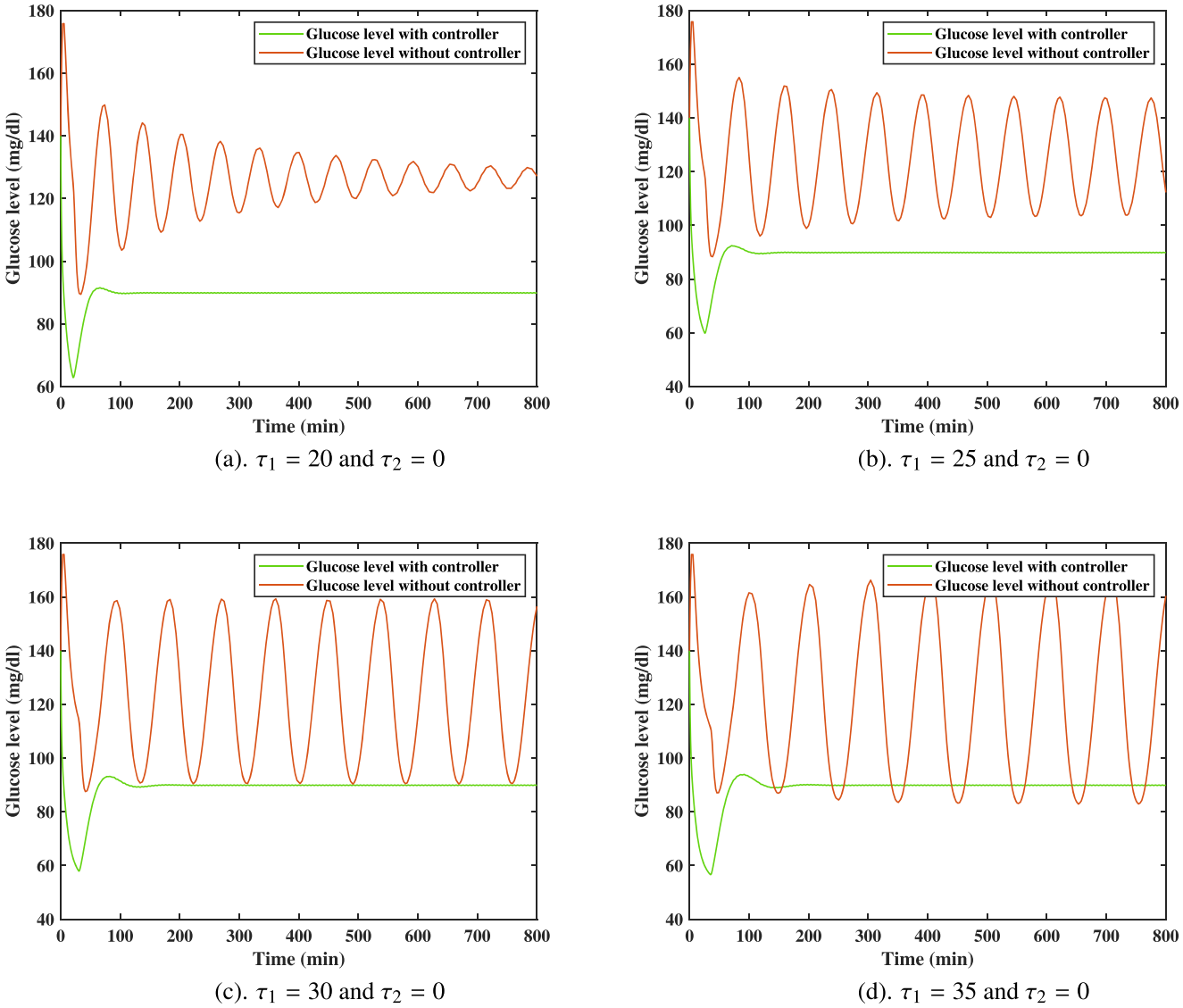
6.1 Glucose profile of patients without meal disturbance

In this type of test, we assumed that the patient had not taken any meal, and controllers examined the patient's blood glucose level response. The simulation work was performed for 800 minutes. The initial concentration of glucose level in a patient was 140 mg/dl. The FL controller maintained the glucose concentration at values 90.4 mg/dl respectively for diabetic patients two hours later.

Figure 5 shows the response of the glucose concentration of patients in which the pancreas delayed the secretion of insulin. The values of τ_1 were taken from 20 to 35 minutes in 5-minute intervals.

Table 4 Parameters of model

Parameters	Units	Values
R_m	$mUmin^{-1}$	210
U_m	$mgmin^{-1}$	940
U_0	$mgmin^{-1}$	40
U_b	$mgmin^{-1}$	72
C_1	mgl^{-1}	2000
C_2	mgl^{-1}	140
C_3	mgl^{-1}	1000
C_4	mgl^{-1}	80
V_i	l	11
V_g	l	10
t_i	min	100
E	$lmin^{-1}$	0.2
a_1	mgl^{-1}	300
β	'unit less'	1.77



(a). $\tau_1 = 20$ and $\tau_2 = 0$

(b). $\tau_1 = 25$ and $\tau_2 = 0$

(c). $\tau_1 = 30$ and $\tau_2 = 0$

(d). $\tau_1 = 35$ and $\tau_2 = 0$

Fig. 5 Glucose concentration for different values of τ_1

Figure 6 represents the glucose levels when the delay occurs in insulin action and insulin secretion in diabetic people. The value of τ_2 was fixed, which was 50 minutes, while the τ_1 value was taken from 10 to 25 minutes in a 5-minute lag.

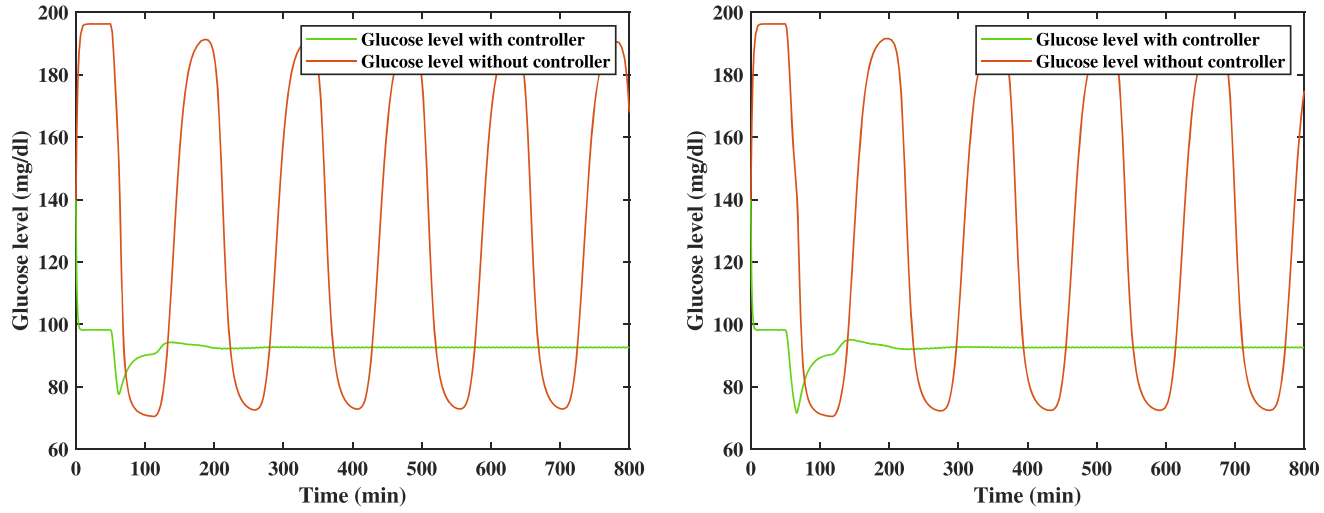
6.2 Glucose profile of patients with meal disturbance

Here, we test how the controllers can deal with the disturbance of multi-mixed meals in diabetic patients. When the patients eat food, its breakdown into glucose molecules and enters the body. This glucose intake rate was assumed a time-dependent function denoted by G_{in}. The maximum glucose intake was at 5.25 mg/min attained at the 15 min mark. The whole duration of the glucose intake lasted 35 min. The proposed controller maintained the glucose level near 90 mg/dl.

$$G_{in} = \begin{cases} 0.25 + \frac{5}{15}t & 0 \le t \le 15(min) \\ 0.25 + 5\frac{35-t}{35-15} & 15 \le t \le 45 \\ 0.25 & 35 \le t \le 240 \end{cases}$$

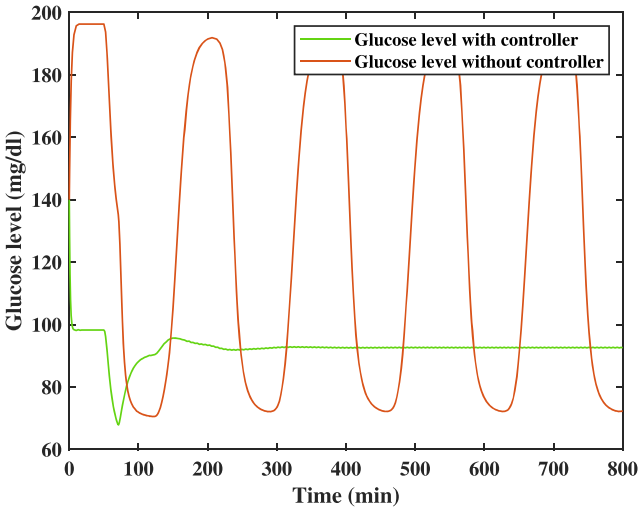
Figure 7 represents the glucose level after the meal announcement in the patient when the delay occurs in the insulin secretion. Here we took the minimum and maximum time delay from the set of τ_1 that was 20 and 35 in the case of $\tau_2 = 0$.

Figure 8 represents the glucose level of patients when the time delay occurs insulin secretion and insulin action

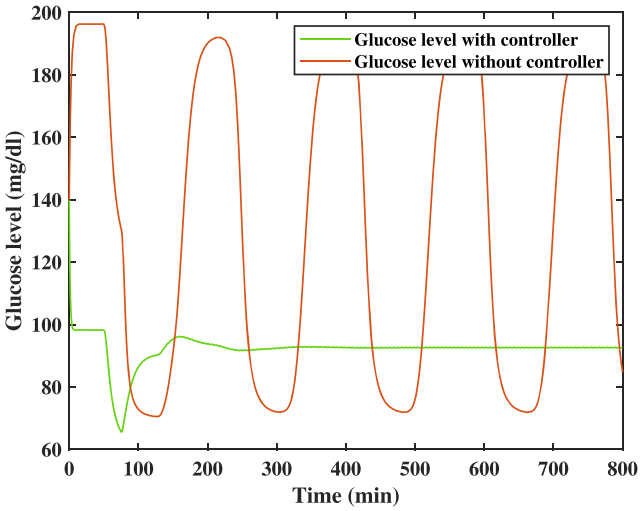


(a). $\tau_1 = 10$ and $\tau_2 = 50$

(b). $\tau_1 = 15$ and $\tau_2 = 50$



(c). $\tau_1 = 20$ and $\tau_2 = 50$



(d). $\tau_1 = 25$ and $\tau_2 = 50$

Fig. 6 Glucose concentration for $\tau_2 = 50$ and different values of τ_1

simultaneously after the meal intake. The minimum and maximum values of the set of τ_1 were 10 and 25, where τ_2 was equal to 50 min.

6.3 Unusual glucose intake by patients

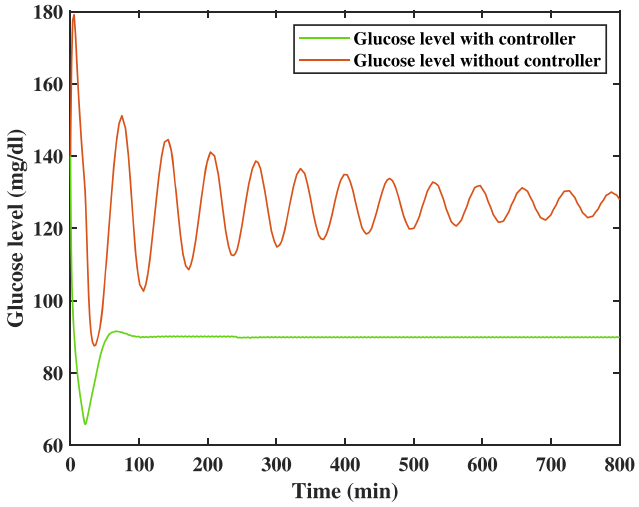
In this protocol, the highest glucose intake rate of the patient was 5 (mg/dL) per minute. At a time interval between 400 and 410 min, the unusual glucose rate was 5 (mg/dL) per minute. The net glucose intake rate was reached 15 (mg/dL)/min with this unexpected glucose consumption.

Figure 9 demonstrates the glucose level of diabetic patients when $\tau_1 = 20, 35$ and $\tau_2 = 0$. The highest recorded glucose levels within the time interval of applying the disturbance were 148.8 mg/dl and 187.3. After controller implantation, the peak values of glucose level during the simulation were 101.1 mg/dl and 101.4 mg/dl.

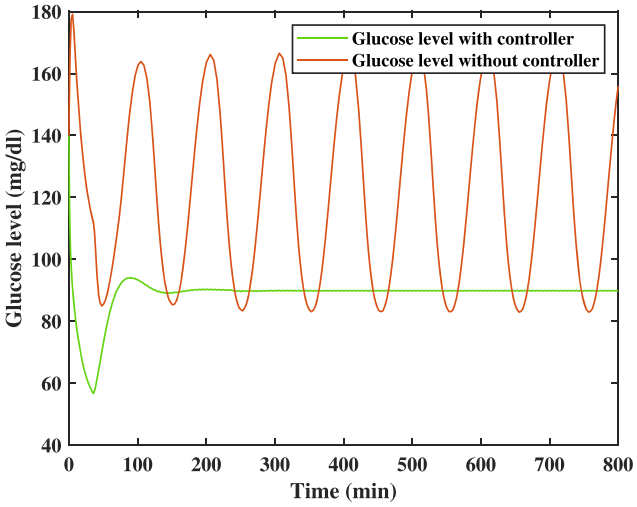
Figure 10 shows the controller's performance in the case of $\tau_1 = 10, 25$ and $\tau_2 = 50$. The highest value of sugar level within the time interval of applying the unusual meal disturbance was 80 mg/dl and 212 mg/dl. After controller implantation, the peak value of glucose level during 800 minutes was 101.1 mg/dl and 103.9 mg/dl.

6.4 Uncertainty in the clearance rate parameter

The insulin clearance rate (d) shows insulin absorption rate, which varies from patient to patient. The high value of this parameter means that the insulin degradation rate increases. Consequently glucose level becomes high in



(a). $\tau_1 = 20$ and $\tau_2 = 0$



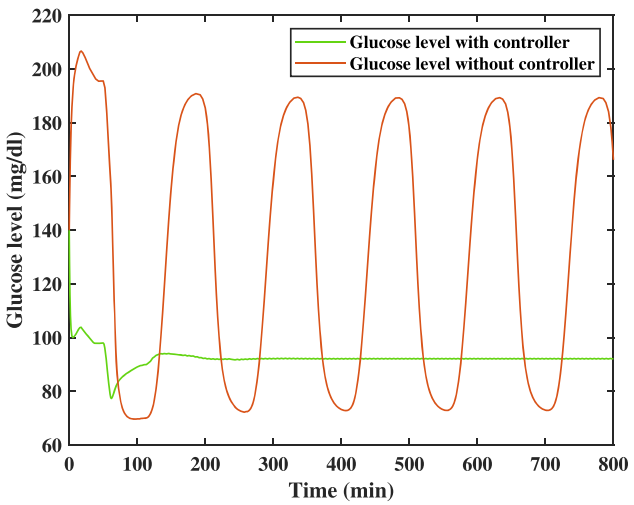
(b). $\tau_1 = 35$ and $\tau_2 = 0$

Fig. 7 Glucose concentration for values of $\tau_1 = 20, 35$ and $\tau_2 = 0$

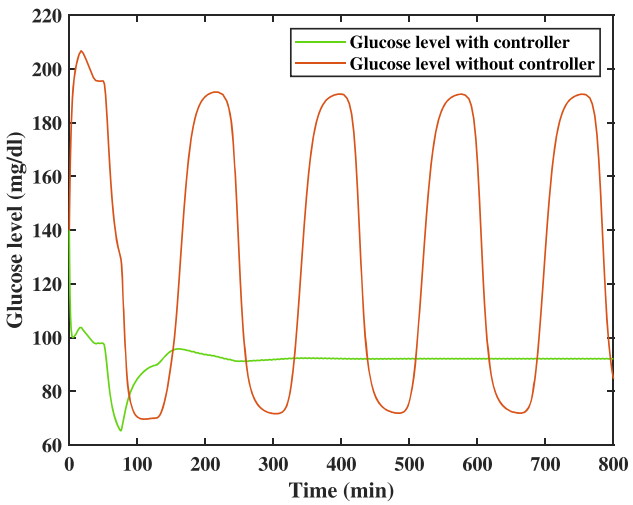
diabetic patients. In this part, simulation was performed based on existing differences between patient regarding their insulin absorption abilty. The parameter of clearance rate considered to be uncertain, whereas its variation was taken to be up to 20%. The nominal value of the parameter was regarded as 0.08 in the model. The results are illustrated in Figs. 11 and 12. The proposed fuzzy logic controller preserverd the glucose concentration in the allowable range.

Figure 11 shows that the controller maintained the glucose level to be in the range of [90.37, 93.01] mg/dL and [90.77, 93.54] mg/dl, after 105 minutes in those diabetic patients who have secreted the 20 min delay insulin, respectively even where $\pm 20\%$ uncertainty in d.

Figure 12 demonstrates that glucose levels in the range of [90.72, 93.54] and [90.92, 93.86] for the diabetic patient with the delay in insulin action were 10 and 25 min, respectively, and insulin absorption that was 50 min.

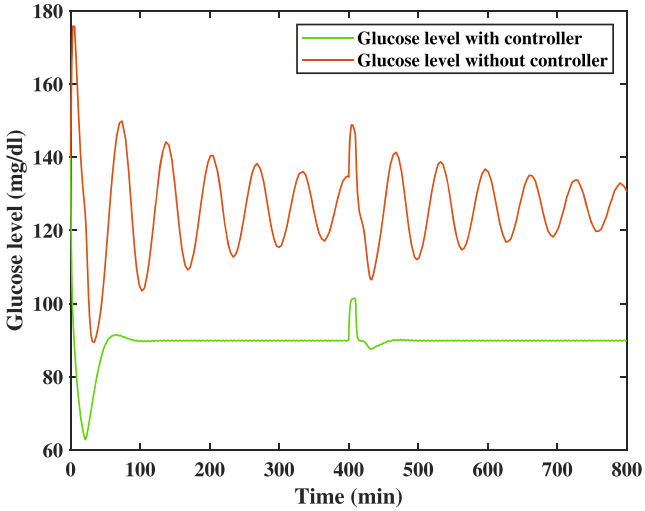


(a). $\tau_1 = 10$ and $\tau_2 = 50$

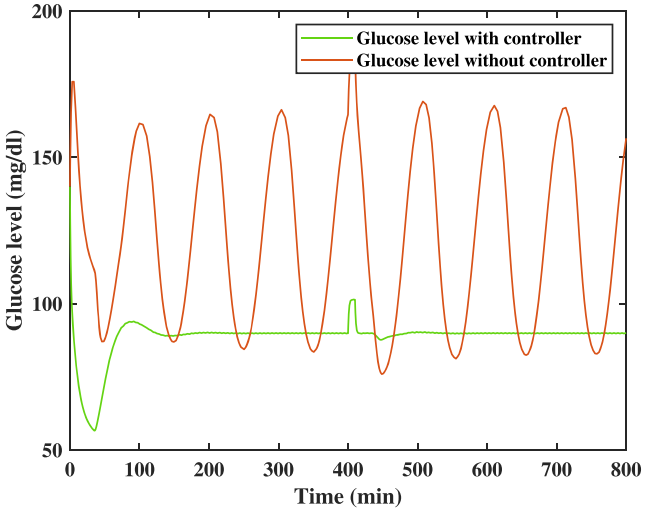


(b). $\tau_1 = 25$ and $\tau_2 = 50$

Fig. 8 Glucose concentration for values of $\tau_1 = 10, 25$ and $\tau_2 = 50$

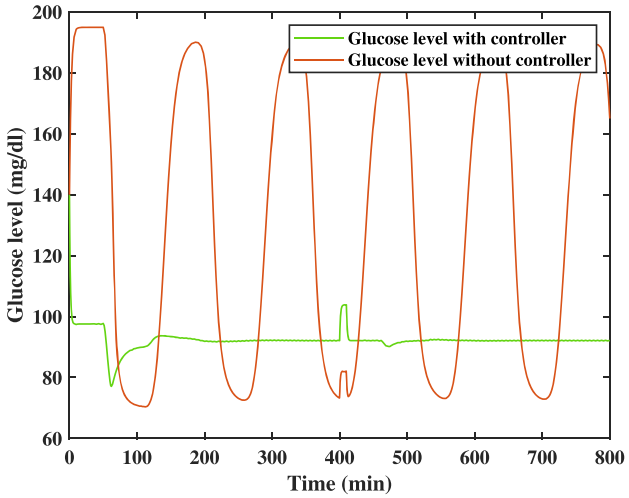


(a). $\tau_1 = 20$ and $\tau_2 = 0$

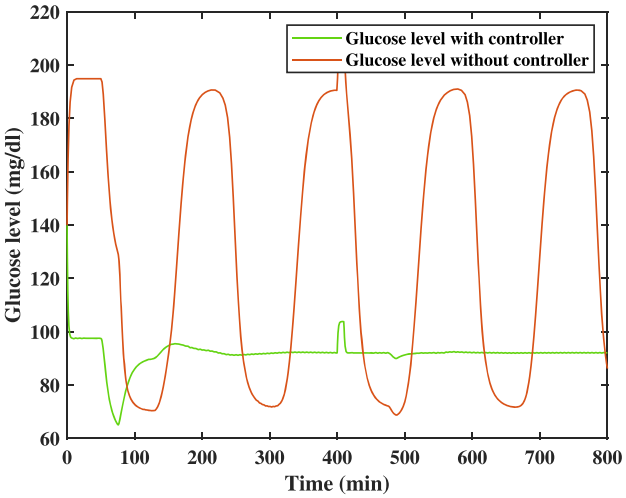


(b). $\tau_1 = 35$ and $\tau_2 = 0$

Fig. 9 Glucose concentration for values of $\tau_1 = 20, 35$ and $\tau_2 = 0$

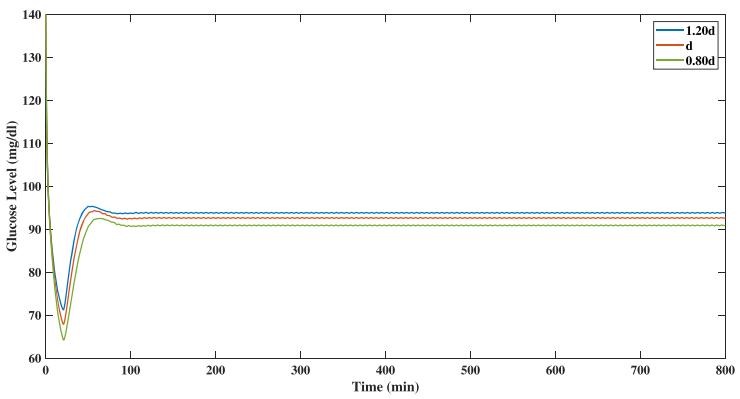


(a). $\tau_1 = 10$ and $\tau_2 = 50$

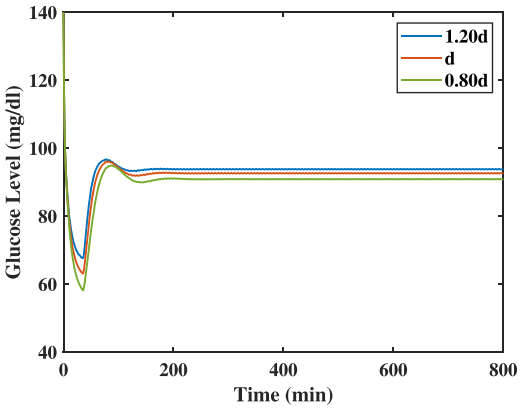


(b). $\tau_1 = 25$ and $\tau_2 = 50$

Fig. 10 Glucose concentration for values of $\tau_1 = 10, 25$ and $\tau_2 = 50$

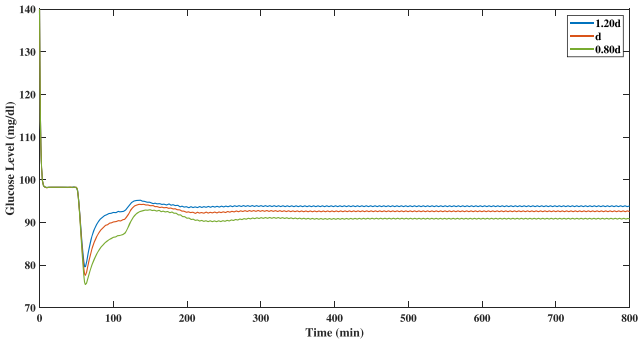


(a). $\tau_1 = 20$ and $\tau_2 = 0$

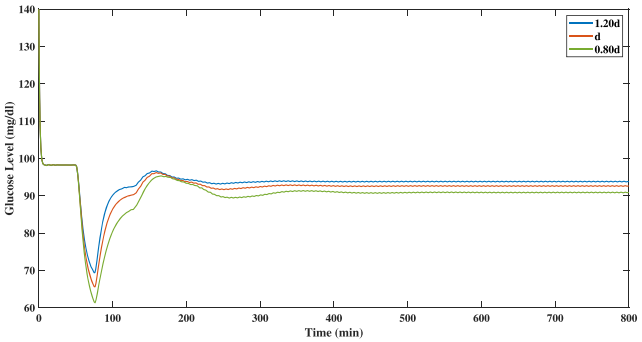


(b). $\tau_1 = 35$ and $\tau_2 = 0$

Fig. 11 Glucose concentration for values of $\tau_1 = 20, 35$ and $\tau_2 = 0$



(a). $\tau_1 = 10$ and $\tau_2 = 50$



(b). $\tau_1 = 25$ and $\tau_2 = 50$

Fig. 12 Glucose concentration for values of $\tau_1 = 10, 25$ and $\tau_2 = 50$

7 Conclusion and future work

In this paper, a two delay non-linear delay differential model was used for describing the glucose-insulin regulatory system. This paper aimed to develop a commendable controller, which helps maintain the glucose profile similar to healthy people in diabetic patients. For this purpose, we proposed an innovative closed-loop controller based on the Mamdani-type fuzzy scheme for type 1 diabetic patients. Generally, it is seen that the delay in insulin secretion and insulin absorption by the body is the pivotal factor of malfunctioning in the blood regulatory system. In the case of a 20 to 35 minutes delay in insulin secretion and a 50-minute delay in insulin absorption, the proposed FLC could keep the glucose levels for the type-I diabetic person in a safe range. The controller was robust, which was proven by the performance of four tests. The controller's ability was evaluated against a high glucose level, which is approximately 2-times greater than the normal glucose level in every experiment. The proposed controller is highly feasible for its use in developing or advancing artificial pancreas, like the micro-pump of insulin infusion. Moreover, the proposed FLC can be easily modified later if some modifications are being carried out in the model. The model's modified factors may depend upon various patient differences, like varying dietary habits, ages, and weights. The type-2 fuzzy controller may be superior in performance than the type-1 fuzzy for dealing with disturbances and uncertainty in model parameters. We recommend using the controller based on the type-2 fuzzy logic technique in the future.

Acknowledgements The authors express their gratitude to the handling editor and anonymous reviewers for their valuable suggestions and comments that improve the manuscript. We are also grateful to the Delhi Technological University for providing financial supports for this research.

Declarations

Conflict of Interest The authors declare that they have no conflict of interest to disclose.

References

1. Nilam, Rathee S (2015) Quantitative analysis of time delays of glucose - insulin dynamics using artificial pancreas. Discret Contin Dyn Syst - Ser B 20(9):3115–3129
2. Esna-Ashari M, Zekri M, Askari M, Khalili N (2017) Predictive Control of the Blood Glucose Level in Type i Diabetic Patient Using Delay Differential Equation Wang Model. J Med Signals Sens 7(1):8–20
3. Soylu S, Danisman K (2018) Blood glucose control using an ABC algorithm-based fuzzy-PID controller. Turk J Electr Eng Comput Sci 26(1):172–183
4. Wang H, Li J, Kuang Y (2009) Enhanced modelling of the glucose-insulin system and its applications in insulin therapies. J Biol Dyn 3(1):22–38
5. Al Gizi A, Mustafa M, A. Alsaedi M, Zreen N (2013) Fuzzy control system review. Int J Sci Eng Res 4(1):1–8
6. Mauseth R, Hirsch IB, Bollyky J, Kircher R, Matheson D, Sanda S, Greenbaum C (2013) Use of a "fuzzy logic" controller in a closed-loop artificial pancreas. Diabetes Technol Therapeut 15(8):628–633
7. Grant P (2007) A new approach to diabetic control: Fuzzy logic and insulin pump technology. Med Eng Phys 29(7):824–827
8. Ibbini MS, Masadeh MA (2005) A fuzzy logic based closed-loop control system for blood glucose level regulation in diabetics. J Med Eng Technol 29(2):64–69
9. Kardar L, Fallah A, Gharibzadeh S, Moztarzadeh F (2008) Application of fuzzy logic controller for intensive insulin therapy in type 1 diabetic mellitus patients by subcutaneous route
10. Liu SW, Huang HP, Lin CH, Chien IL (2013) Fuzzy-logic-based supervisor of insulin bolus delivery for patients with type 1 diabetes mellitus. Ind Eng Chem Res 52(4):1678–1690
11. Mollaei N, Kardehi Moghaddam R (2014) A New Controlling Approach of Type 1 Diabetics Based on Interval Type-2 Fuzzy Controller
12. Goharimanesh M, Akbari AA (2015) A Comparison of Fuzzy Types 1 and 2 in Diabetics Control, Based on Augmented Minimal Model. J World's Electr Eng Technol 4(2):70–75

13. Gharghory SM, El-Dib DA (2016) Fuzzy Control System for Regulating the Blood Glucose Level of Diabetes Patients Implemented on FPGA. J Circ Syst Comput 25(12):1650161
14. Amuthameena S (2016) A novel strategy for blood glucose control in human body using PID-Fuzzy Logic Controller. J Chem Pharmaceut Sci 8, pp 88–92
15. Chen J, Cao K, Sun Y, Xiao Y, Su X (2007) Continuous drug infusion for diabetes therapy: A closed-loop control system design. Eurasip J Wirel Commun Netw 2008:1–10
16. Journal OA, Nagapavithra S (2017) Diabetes Type -1 Insulin level Control using closed loop control of BLDC motor. Adv Natural Appl Sci 11(6):728–734
17. Sasi AYB, Elmalki M (2013) A Fuzzy Controller for Blood Glucose-Insulin System. J Signal Inf Process 04(02):111–117
18. Allam F, Nossair Z, Gomma H, Ibrahim I, Abdelsalam M (2012) Evaluation of Using a Recurrent Neural Network (RNN) and a Fuzzy Logic Controller (FLC) In Closed Loop System to Regulate Blood Glucose for Type-1 Diabetic Patients. Int J Intell Syst Appl 4(10):58–71
19. Khooban MH, Maryam ABADI DN, Alfi A, Siahi M (2013) Swarm optimization tuned mamdani fuzzy controller for diabetes delayed model. Turk J Electr Eng Comput Sci 21(SUPPL. 1):2110–2126
20. Abadi DNM, Khooban MH, Alfi A, Siahi M (2014) Design of Optimal Self-Regulation Mamdani-Type Fuzzy Inference Controller for Type I Diabetes Mellitus. Arab J Sci Eng 39(2):977–986
21. Jain V, Raheja S (2015) Improving the Prediction Rate of Diabetes using Fuzzy Expert System. Int J Inf Technol Comput Sci 7(10):84–91
22. Reddy GT, Khare N (2016) FFBAT- Optimized rule based fuzzy logic classifier for diabetes. Int J Eng Res Afr 24(June):137–152
23. Gaz C, De Gaetano A, Manes C, Palumbo P, Borri A, Panunzi S (2017) Effective control of glycemia using a simple discrete-delay model IFAC-PapersOnLine, vol 50. Elsevier B.V., pp 13526–13531
24. Engelborghs K, Lemaire V, Bélair J, Roose D (2001) Numerical bifurcation analysis of delay differential equations arising from physiological modeling. J Math Biol 42(4):361–385
25. Sturis J, Polonsky KS, Mosekilde E, Van Cauter E (1991) Computer model for mechanisms underlying ultradian oscillations of insulin and glucose. Amer J Physiol-Endocrinol Metabolism 260(5):E801–E809
26. Li J, Kuang Y, Mason CC (2006) Modeling the glucose-insulin regulatory system and ultradian insulin secretory oscillations with two explicit time delays. J Theor Biol 242(3):722–735
27. Iva TM, Mosekilde ERIKM, Turis JES (2000) Modeling the InsulinGlucose Feedback System: The Significance of Pulsatile Insulin Secretion. J Theor Bil 207:361–375
28. Panunzi S, Palumbo P, De Gaetano A (2007) A discrete Single Delay Model for the Intra-Venous Glucose Tolerance Test. Theor Biol Med Model 4:1–16
29. Wang H, Li J, Kuang Y (2007) Mathematical modeling and qualitative analysis of insulin therapies. Math Biosci 210(1):17–33
30. Li J, Johnson J (2009) Mathematical models of subcutaneous injection of insulin analogues: A mini-review. Discret Contin Dyn Syst - Ser B 12(2):401–414
31. Makroglou A, Karaoustas I (2011) A review on delay differential equation models in diabetes modeling, II: the insulin therapies and the intracellular activities of β-cells case. Math Comput Simul:1–24
32. Makroglou A, Karaoustas I, Li J, Kuang Y (2009) Delay differential equation models in diabetes modeling : a review. Theor Biol Med Model:1–19
33. Shampine LF, Thompson S, Chalishajar D, Geary DH, Cox G (2016) Review study of detection of diabetes models through delay differential equations. 37, pp 1087–1102
34. Huang M, Li J, Song X, Guo H (2012) Modeling Impulsive Injections of Insulin: Towards Artificial Pancreas. SIAM J Appl Math 72(5):1524–1548
35. Shaikhet L (2016) Stability of equilibrium states of a nonlinear delay differential equation with stochastic perturbations. Int J Robust Nonlinea 27(6):915–924

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



Ankit Sharma is a PhD student at Delhi Technological University's Department of Applied Mathematics. He received the gold medal in the master's program. His current research interests include nonlinear dynamics, stability analysis, and fuzzy logic controller.



Dr. Nilam is faculty in the department of Applied mathematics of Delhi Technological University since Feb 2011. She is doctorate from IIT Roorkee. Her research interest is Mathematical modelling and Simulation, Differential equation, Dynamics of Infectious disease, Delay differential equation, Fractional order differential equations, Control theory in Diabetes. She has vast experience of more than 15 years in teaching and research.



Harendra Pal Singh is the faculty in Cluster Innovation Centre (University of Delhi) since July 2015. He is doctorate from IIT Roorkee. His research interest is Modelling and simulation of engineering systems, Stability analysis of dynamical systems, Machine learning (Robot learning), Fractional order linear and nonlinear systems and controls.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360484597

Deep Learning Methods for Anomaly Detection

Conference Paper · May 2022

CITATIONS
0

READS
49

4 authors, including:



Rajeev Subedi
Dehradun Institute of Technology
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Anil Kumar
DIT University
107 PUBLICATIONS **316** CITATIONS
SEE PROFILE



Narendra Kumar
Dehradun Institute of Technology
31 PUBLICATIONS **141** CITATIONS
SEE PROFILE

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

Optimized Deep Learning Implementation for Improving the Performance of Upstream Petroleum Operations View project

All content following this page was uploaded by Narendra Kumar on 10 May 2022.

The user has requested enhancement of the downloaded file.

Deep Learning Methods for Anomaly Detection

Rajeev Subedi[1, a)] **Anil Kumar** [2, b)] **Narendra Kumar** [3, c)]

Author Affiliations
[1, 2, 3] *School of Computing, DIT University Dehradun, Uttarakhand, India*
Author Emails
[a)] *Corresponding author: 1000014411@dit.edu.in*
[b)]*dahiyaanil@yahoo.com*

[c)]*narendra298@gmail.com*

Abstract.

The invention and upgrading of systems and instruments that collect data at each point are narrowing the world. As a result, data is growing at an exponential rate. We are unable to process such a vast volume of information in a timely manner. The data comprises numbers that change over time, as well as other factors and functions that cause it to fluctuate suddenly. Anomaly is a term used to describe an abrupt change in data from its standard. The purpose of this survey is to present the deep learning approaches and the deep learning architectures with which we can detect anomalies. The review also presents the summarized view of the various approaches, techniques, datasets used to detect the anomalies in the studies done over the years. The prime goal of this paper is to summarize the various deep learning approach, architecture, and datasets with which we can detect anomalies.

Keywords. Anomaly, Outliers, Neural network, Deep learning, Auto encoder.

1. INTRODUCTION

The Anomalies are rapid changes in data from normal deviation into clusters. Anomaly detection is the process of analysing the data provided by machines or sets of machines to find abnormalities. Anomaly detection has recently had a big influence in areas including hospital monitoring systems, security, banking, IoT and sensor networks, marketing, and natural disasters. As a consequence, anomaly detection has been a popular research topic for decades, owing to its intrinsic complexity and tumultuous nature[1]. While traditional techniques of identifying anomalies have existed for a long time, deep learning emerged to outwit numerous learning tasks by learning complicated data using neural networks, which are computer representations of brain neurons. Many research methodologies are employed to build the most effective model that can identify irregularities with a reduced signal to noise ratio. The goal of such a model is to detect as many outlier clusters as possible while reducing risk. The necessity for new models that can evaluate enormous datasets was essential since previous approaches could not handle such quantities to create outliers. Traditional approaches failed to optimise time-series, picture, and sequential data due to their complicated structures[2]. Whereas a standard model relies on human feature selection

from a dataset, a model using deep learning outlier detection may learn graded discriminative features[2].

2. DIFFERENT DEEP LEARNING APPROACHES FOR ANOMALY DETECTION

2.1. Supervised anomaly detection

Separating training and test datasets is essential for supervised anomaly detection. Despite better efficiency and outcomes, semi-supervised and unsupervised techniques of anomaly detection were more reliable due to greater unlabeled training samples in the created dataset [2][3].

2.2. Unsupervised Anomaly Detection

Unsupervised deep learning anomalydetection builds and trains a model using an unclassified and uncategorized dataset[1]. A model evaluates unscaled datasets and predicts hidden patterns. It is explained in the study publication [2] that an unsupervised anomaly detection system is able to learn from the data and find anomalies by separating normal from abnormal data points.

2.3. Semi-Supervised Anomaly Detection

It is using both scaled and unscaled datasets to train a model is called hybrid learning[4][5].The scaled to unscaled data ratio is decreasing, indicating that less scaled data is used for training and more unscaled data is used for better outcomes [6].

3. DIFFERENT ARCHITECTURES IN DEEP LEARNING FOR ANOMALY DETECTION

3.1. Multi-layer Perceptron neural network

Multilayer perceptron (MLP) is a multiple layered feed-forward neural network[7], [8].

These activation functions[9] are represented as

$$y(v_i) = \tanh(v_i) \quad (3.1)$$

$$y(v_i) = (1 + e^{-vi})^{-1} \quad (3.2)$$

Here, back-propagation, popular supervised learning technique is used for training the dataset[9].

The multilayer perceptron neural network as given:

Let us consider d_j as the targeted value and y_j as the output value calculated by the perceptron algorithm. Now the error in one neuron is generated by:

$$e_j(n) = d_j(n) - y_j(n) \quad (3.3)$$

with which our error function is calculated by:

$$\varepsilon(n) = \frac{1}{2} \Sigma e_j^2(n) \quad (3.4)$$

Now apply the gradient descent, the value while updating the weights in each cycle is calculated by:

$$\Delta w_{ij}(n) = -\eta (d\varepsilon / dv_j) y_i(n) \quad (3.5)$$

3.2. Convolution Neural Network(CNN)

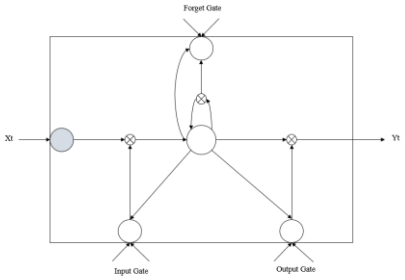
CNN prioritises the local dependencies in the input data gathered earlier in the process. Many engineers, academicians built CNN models using image datasets. A variety of methodologies and datasets have been used by other researchers to create the model [10]-[11].

3.3. Recurrent Neural Network

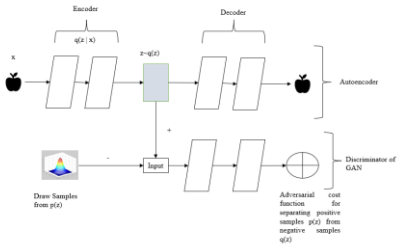
It is a type of advanced deep learning that operates by sending the hidden layer output back to the input layer, resulting in the prediction of that layer's output [14]. Unlike other neural networks, RNNs have an internal memory that allows them to perform the same function for each input.

3.4. Long Short-Term Memory

In Researchers developed the LSTM as a sophisticated RNN in the late 1990s and early 2000s[10]. The vanishing gradient issue in RNN happens when back-propagation is used to train a model [17]. LSTM is an efficient solution. Advanced recurrent neural network architecture that uses previously stored data to analyse and forecast abnormalities [15].



(a) LSTM network architecture[14]



(b) An architecture of adversarial autoencoder[19]

Figure 3.1 An architecture of LSTM Network and adversarial autoencoder

3.5 Adversarial Auto Encoder

Adversarial Autoencoder (AAE) is a promising strategy in deep learning to identify anomalies that may transform an autoencoder model into a generative adversarial network with the primary goal of minimising AE reconstruction error and mapping a prior to the hidden code vector[13].

3.6 Restricted Boltzmann Machine

It is a stochastic neural network controlled by energy principles and distinguished by variables such as entropy, energy, and temperature[16]. With applications ranging from feature extraction to filtering to pre-training neural networks to picture reconstruction [33-42], networks are increasingly becoming unsupervised learning machines. RBM is an

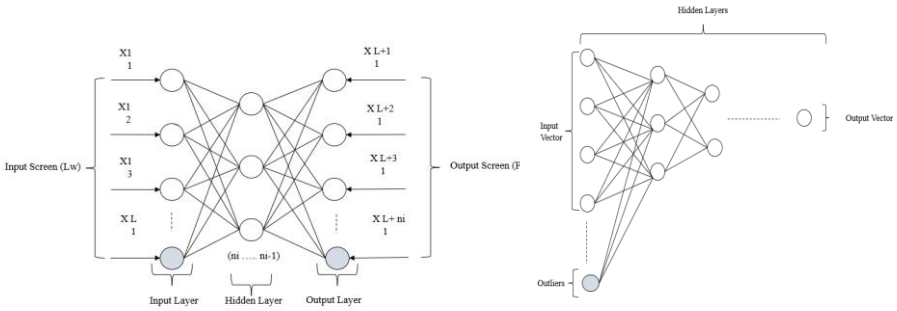
unsupervised learning technique with two layers: input and hidden. Because there is no output layer, the propagation entails returning the reconstruction to the input layer[15].

3.7 Generative Adversarial Networks

In order to simulate the large variety of distribution of complicated and multi-dimensional data, we need more sophisticated approaches, such as generative adversarial networks (GANs)[16],[17]. The primary flaw of GANs is that they are unstable to train and provide absurd outputs.

3.8 Deep Belief Networks

Deep Belief networks (DBNs)[18] are generative model for neural network that stack RBMs layer by layer and carry out the learning process.



(a) Multilayer perceptron network architecture[7] **(b)**Deep belief network architecture

Figure 3.3 Multilayer perceptron and Deep belief network architecture[21].

Above architecture of DBN shows us the combine structure of simple and hierarchical connected RBMs [19].

3.9 Autoencoder

The key objective of an autoencoder is to learn in an unsupervised manner interpretation of the dataset which can later be used to solve several real-life applications[20].Also, several variations of regularization techniques for autoencoders were studied in various case studies which includes Sparse Autoencoders[21]–[23],Denoising Autoencoders[24], [25] and Contractive Autoencoders[26].

3.10 Deep Neural Networks

Moreover, with these features DNNs is widely used for anomaly detection methods surpassing the effectiveness and accuracies of MLP, CNN, RNN, LSTM, RBM and LSTM and are being used as black boxes in a model with their dense non-linear structure[27], [28].

4. COMPARISON RESULTS

Supervised, semi-supervised and unsupervised approaches were presented in the paper reviewed throughout the research, but most importantly unsupervised learning approach is most desired for anomaly detection.Couple of more anomaly detection architecture based on LSTM[29], LSTM-SVM[30], CNN-LSTM[11], LSTM-RNN[32] and Stacked

LSTM[31] were observed where we could see various decision function and some methods are extremely good where maximum ratio of anomalies was detected whereas some were average.This research was conducted by making model to detect anomalies from normal patters using deep Auto-encoder and restricted Boltzmann machine which uses backpropagation by setting equal inputs, outputs and generate the results based on MSE, RMSE and AUC.Last but not the least, we observed Deep Auto-encoder (DAE) architecture [33]which elaborates the need of hybrid semi-supervised anomaly detection model which could minimize the complexity observed in dimensionality in high dimensional space for high-dimensional data.

5. CONCLUSION

This paper presents the Deep learning anomaly detection approach and methods along with the literature survey based on the approach, detection technique, architecture, datasets used, and result obtained. The prime goal of this paper was to summarize the various deep learning approach, architecture, and datasets with which we can detect anomalies. Another aspect of this research was to find the application areas within deep learning for which we can opt anomaly detection technique. Anomaly detection using deep learning technique is the major area of research and this paper could be useful for researchers in the field.

6. REFERENCES

1. G. Pang, C. Shen, L. Cao, and A. van den Hengel, ACM Computing Surveys 54, (2021).
2. R. Chalapathy and S. Chawla, (2019).
3. N. Görnitz NICOGOERNITZ, K. Rieck KONRADRIECK, and U. Brefeld, Toward Supervised Anomaly Detection Marius Kloft (2013).
4. B.R. Kiran, D.M. Thomas, and R. Parakkal, Journal of Imaging 4, (2018).
5. E. Min, J. Long, Q. Liu, J. Cui, Z. Cai, and J. Ma, in Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics) (Springer Verlag, 2018), pp. 322–334.
6. P. Perera and V.M. Patel, IEEE Transactions on Image Processing 28, 5450 (2019).
7. G. Raman MR, N. Somu, and A.P. Mathur, International Journal of Critical Infrastructure Protection 31, (2020).
8. R. Collobert and S. Bengio, Links between Perceptrons, MLPs and SVMs (2004).
9. 9 T. Teoh, P. Ng, G. Chiew, E.J. Franco, and de Y. Goh, Anomaly Detection in Cyber Security Attacks on Networks Using MLP Deep Learning (n.d.).
10. A. Krizhevsky, I. Sutskever, and G.E. Hinton, Communications of the ACM 60, 84 (2017).
11. Y. Heryadi and H.L.H.S. Warnars, in 2017 IEEE International Conference on Cybernetics and Computational Intelligence, CyberneticsCOM 2017 - Proceedings (Institute of Electrical and Electronics Engineers Inc., 2018), pp. 84–89.
12. A. Dimokranitou, G. Tsechpenakis, J. Yu Zheng, and M. Tuceryan, STATEMENT OF THESIS APPROVAL (2017).
13. G.H. de Rosa, M. Roder, D.F.S. Santos, and K.A.P. Costa, International Journal of Information Technology (Singapore) 13, 49 (2021).
14. A. Pumsirirat and L. Yan, Credit Card Fraud Detection Using Deep Learning Based on Auto-Encoder and Restricted Boltzmann Machine (2018).

15. I.J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, (2014).
16. A. Radford, L. Metz, and S. Chintala, (2015).
17. M.K. Sharma, D. Sheet, and P.K. Biswas, in ACM International Conference Proceeding Series (Association for Computing Machinery, 2016).
18. I. Kakanakova and S. Stoyanov, in ACM International Conference Proceeding Series (Association for Computing Machinery, 2017), pp. 73–79.
19. D. Bank, N. Koenigstein, and R. Giryes, (2020).
20. P. Malhotra, L. Vig, N. Gugulothu, and G. Shroff, Sparse Neural Networks for Anomaly Detection in High-Dimensional Time Series (2018).
21. M.R. Shahid, G. Blanc, Z. Zhang, and H. Debar, (2019).
22. J. Sun, X. Wang, N. Xiong, and J. Shao, IEEE Access 6, 33353 (2018).
23. J. Chen, J. Li, W. Chen, Y. Wang, and T. Jiang, Renewable Energy 147, 1469 (2020).
24. M.G. Narasimhan and S. Sowmya Kamath, Multimedia Tools and Applications 77, 13173 (2018).
25. S.F. Lokman, A.T. Othman, S. Musa, and M.H. Abu Bakar, in Advanced Structured Materials (Springer Verlag, 2019), pp. 195–205.
26. K. Amarasinghe, K. Kenney, and M. Manic, in Proceedings - 2018 11th International Conference on Human System Interaction, HSI 2018 (Institute of Electrical and Electronics Engineers Inc., 2018), pp. 311–317.
27. W. Samek, T. Wiegand, and K.-R. Müller, (2017).
28. G. Kim, H. Yi, J. Lee, Y. Paek, and S. Yoon, (2016).
29. T. Ergen and S.S. Kozat, IEEE Transactions on Neural Networks and Learning Systems 31, 3127 (2020).
30. P. Baldi, Autoencoders, Unsupervised Learning, and Deep Architectures (2012).
31. T.K. Dang, R. Wagner, J. Küng, N. Thoai, M. Takizawa, and E. Neuhold, editors , Future Data and Security Engineering (Springer International Publishing, Cham, 2016).
32. H. Song, Z. Jiang, A. Men, and B. Yang, Computational Intelligence and Neuroscience 2017, (2017).
33. N. Kumar, H. Shukla and R. Tripathi, International Journal Of Intelligent Engineering And Systems 10, (2017).
34. H. Shukla, N. Kumar and R. Tripathi, International Journal Of Computer Applications 95, (2014).
35. N. Kumar, A. Dahiya, K. Kumar and S. Tanwar, 2021 9Th International Conference On Reliability, Infocom Technologies And Optimization (Trends And Future Directions) (ICRITO) (2021).
36. K. Kumar, N. Kumar and R. Shah, International Journal Of Intelligent Networks 1, (2020).
37. N. Kumar, H. Shukla and R. Tripathi, International Journal Of Intelligent Engineering And Systems 10, (2017).
38. A. Sharma, N. Kumar, International Journal of Science and Research (IJSR) 3,(2014)
39. N. Kumar, H. Shukla, A. Tiwari and A. Dahiya, SSRN Electronic Journal (2019).
40. N. Kumar, H. Shukla and R. Tripathi, International Journal Of Intelligent Engineering And Systems 10, (2017).
41. N. Kumar, K. Kumar and A. Kumar, Sersc.Org (2020).
42. K. Kumar, R. Singh, P. Ranjan and N. Kumar, Algorithms For Intelligent Systems ((2021).

Biographies



Er. Rajeev Subedi received his bachelor's degree (B.Tech) in Computer Science & Engineering from Chandigarh University in 2019 and currently perusing M.Tech in Computer Science & Engineering from DIT University respectively. He is also working as an IT consultant in multiple schools and colleges in Nepalgunj, Nepal. His research areas include machine learning, deep learning, and cloud computing.



Dr. Anil Kumar is now employed at DIT University as a Professor of CSE,Head-Data Science Research Group and Accreditation Coordinator . Prof. Anil Kumar received his M.Tech. from the Delhi College of Engineering and his Ph.D. from the Manipal Group. He has over 24 years of teaching and working experience. He is an IEEE Senior Member and served on the Executive Committee of the IEEE Computer Society India Council in 2015 and 2016. His research interests include cryptography, image processing algorithms, artificial intelligence, neural systems, signal and system analysis, and genetic algorithms.



Dr. Narendra Kumar completed his M.Tech(Computer Science) from BIT Mesra Ranchi, and Ph.D.(Computer Science) from D.D.U Gorakhpur University. Narendra Kumar has more than 14 years' rich experience in Computer Science & Engineering field. He is working at DIT University as an Assistant Professor for the last 14 years, His research area includes image processing, Optimizations techniques and the internet of things and deep learning. Narendra is Editor of two books including publication like CRC Press, Taylor & Francis Group and Springer publication. He has published numerous research papers in international journals and conferences including Springer, IEEE, and Elsevier.

View publication stats

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Design and Analysis of 55–63-GHz Fundamental Quad-Core VCO With NMOS-Only Stacked Oscillator in 28-nm CMOS

Sriram Balamurali, *Student Member, IEEE*, Giovanni Mangraviti, *Member, IEEE*, Cheng-Hsueh Tsai, *Student Member, IEEE*, Piet Wambacq, *Senior Member, IEEE*, and Jan Craninckx, *Fellow, IEEE*

***Abstract*— This article presents a 55–63-GHz fundamental multicore voltage-controlled oscillator (VCO) in a 28-nm bulk CMOS process. The single-core VCO utilizes stacking and magnetic coupling of two NMOS-based resonators, which increases the tank energy for phase noise reduction while maintaining low voltage swings for reliability. In the multi-core configuration, the oscillators are coupled in a way that also simplifies LO distribution toward the transmitter and receiver. The potential parasitic modes that occur both in single-core and multi-core configurations are suppressed by design. The implicit common-mode (CM) impedance control is utilized to reduce phase noise without any large extra tail inductor as used in a conventional complementary oscillator. The distributed nature of energy restoring elements in the proposed stacked oscillator also offers robust parasitic mode suppression compared to alternative implementation based on a conventional complementary oscillator. The working range of the prototype is between 55 and 63 GHz corresponding to a tuning range of 13.4% with power consumption varying from 15.3 to 12 mW across the tuning range. The prototype achieves the best phase noise of −95 and −121.7 dBc/Hz at 1- and 10-MHz offsets from the carrier at 55GHz. The corresponding peak figure of merits (FOMs) achieved are 178 and 185 dBc/Hz at 1- and 10-MHz offsets from the carrier.**

***Index Terms*— Common-mode (CM) resonance, impulse-sensitivity function (ISF), millimeter-wave voltage-controlled oscillator (VCO), multi-core oscillator, parasitic mode and Kärtner's equations, phase noise, stacked oscillator.**

I. Introduction

EMERGING mm-wave wireless communication systems have stringent LO phase noise specifications to support higher order modulation schemes. Prior works have shown that a low phase noise LO can be synthesized by extracting higher order harmonics of a low-frequency oscillator [1]–[4]. However, much power is spent on harmonic extraction of weak harmonics at mm-wave frequencies. Besides this, the LO has to be distributed to up- and down-conversion mixers of TX and RX. This distribution is one of the most power-hungry operations in millimeter-wave systems. The fundamental multi-core voltage-controlled oscillator (VCO) presented in this article has a layout with four remote ends at which an LO signal can be tapped [5]. In this work, we combine parallel and magnetic couplings of the oscillators to lower the phase noise and to do power-efficient LO distribution.

The tradeoff between the energy stored in the tank (E_{tank}) and the tank quality factor (Q) limits the phase noise achievable in a single-core oscillator. This can be seen from Leeson's equation (1) and its alternative form rewritten as (2) [6]. Here, F is the effective noise factor dependent on the impulse-sensitivity function (ISF), K_B is Boltzmann's constant, T is the temperature, ω_o is the resonance frequency, $\Delta\omega$ is the frequency offset from the carrier, and A_{pk} is the peak differential amplitude. E_{thermal} is the mean thermal energy per degree of freedom

$$L(\Delta\omega) = \frac{4FK_BTR_T}{V_{\text{pk}}^2}\left(\frac{\omega_o}{2Q\Delta\omega}\right)^2 \tag{1}$$

$$L(\Delta\omega) = \frac{FK_BT}{C_TV_{\text{pk}}^2}\cdot\frac{\omega_o}{Q}\cdot\frac{1}{\Delta\omega^2} = \frac{\text{FE}_{\text{thermal}}}{E_{\text{tank}}}\cdot\frac{\omega_o}{Q}\cdot\frac{1}{\Delta\omega^2}. \tag{2}$$

From (2), it is obvious that the phase noise can be reduced by increasing the energy in the resonator. E_{tank} can be written either as peak energy stored in the capacitor, i.e., $1/2CV^2$ or peak energy stored in the inductor, $1/2V^2/L\omega_o^2$. Decrease in the inductor size to increase the tank energy after certain optimum results in the degradation in-tank Q [6], hence this tradeoff. One of the ways for overcoming E_{tank} versus Q tradeoff is by parallel coupling of the oscillators. In principle, parallel coupling N oscillators result in N times more energy and, hence, $10\log N$ times better phase noise if Q is preserved [1], [7]–[12]. In our multi-core oscillator, we also utilize parallel coupling to reduce phase noise and synchronize four single-core stacked oscillators. This makes LO outputs available at four remote ends of the layout.

The peak swing of an NMOS-only oscillator is twice that of a complementary oscillator. Thus, fundamentally, it can store 4× more energy and, thus, by (2), can have 6-dB better phase noise for the same figure of merit (FOM). NMOS-only

Manuscript received December 19, 2021; revised February 24, 2022; accepted March 27, 2022. This article was approved by Associate Editor Salvatore Levantino. This work was supported in part by the Strategic Research Program of the Vrije Universiteit Brussels under Grant SRP-19. *(Corresponding author: Sriram Balamurali.)*

Sriram Balamurali and Piet Wambacq are with imec, 3001 Leuven, Belgium, and also with the Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, Belgium (e-mail: sriram.balamurali@imec.be).

Giovanni Mangraviti and Jan Craninckx are with imec, 3001 Leuven, Belgium.

Cheng-Hsueh Tsai is with the Department of Electronics and Informatics, Vrije Universiteit Brussel, 1050 Brussels, Belgium.

Color versions of one or more figures in this article are available at https://doi.org/10.1109/JSSC.2022.3165654.

Digital Object Identifier 10.1109/JSSC.2022.3165654

0018-9200 © 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 May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

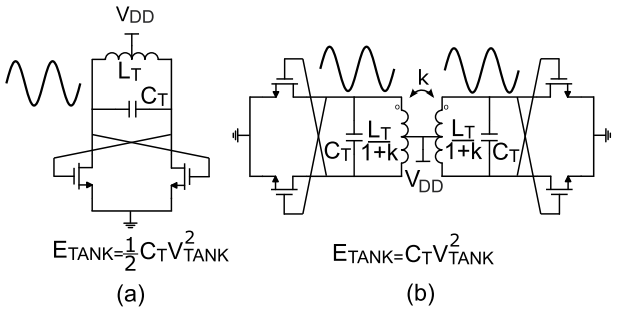


Fig. 1. Magnetic coupling to increase the tank energy.

oscillators operating at nominal supply voltage, however, cannot pass stringent reliability requirements with thin oxide transistors [7]. To still meet reliability requirements using only NMOS devices, thick oxide transistors can be used. However, they drastically impact millimeter-wave performance. Another way is to reduce supply to half the nominal system supply, thus needing high-dropout voltage regulators and reducing oscillator power efficiency. Four coupled complementary oscillators store the same energy as one single NMOS-only oscillator at nominal supply voltage and, hence, overcome these limitations of NMOS-only oscillators working at nominal supply voltage [7]. Our work utilizes NMOS device stacking to overcome the same limitations.

Section II discusses the proposed stacked oscillator. The physics behind the phase noise reduction due to magnetic coupling in the proposed oscillator will be discussed intuitively through the ISF analysis. The proposed stacked oscillator using thin oxide devices is used to increase the tank energy and, thus, reduce phase noise without affecting the device's reliability. Compared to complementary work [7], common-mode (CM) impedance is controlled implicitly without requiring extra big tail inductors. As will be seen later, these extra common-mode inductors not only increase the oscillator area without storing any extra energy but also impede LO distribution. Section III discusses the dual-core version of the proposed oscillator. The associated parasitic modes are studied and discussed. It will be seen that the distributed nature of energy restoring elements in stacked oscillators also leads to robust parasitic mode suppression with a low value of coupling resistor. In a multi-core oscillator, four of these single-core units are coupled for phase noise reduction in a configuration that also simplifies LO distribution as will be seen in Section IV. Sections V and VI will discuss the measurement results of the prototype in 28-nm CMOS and the conclusions of the work.

II. Single-Core Stacked Oscillator

A. Proposed Stacked Oscillator

Magnetic coupling between two VCOs is one of the ways to increase the energy stored in the resonator (see Fig. 1) and, thus, reduce phase noise.

In the proposed single-core VCO [see Fig. 2(b)], two NMOS cross-coupled pairs are stacked within a single supply and magnetically coupled to increase the energy stored in the tank. Stacking is done for reliable nominal supply operation. The primary side of the transformer is used for LO distribution and

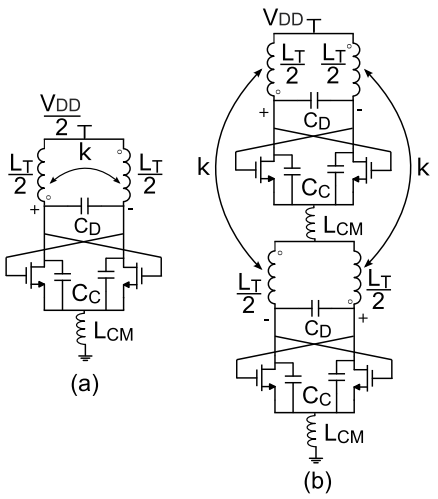


Fig. 2. (a) Exemplary NMOS oscillator. (b) Stacked NMOS oscillator.

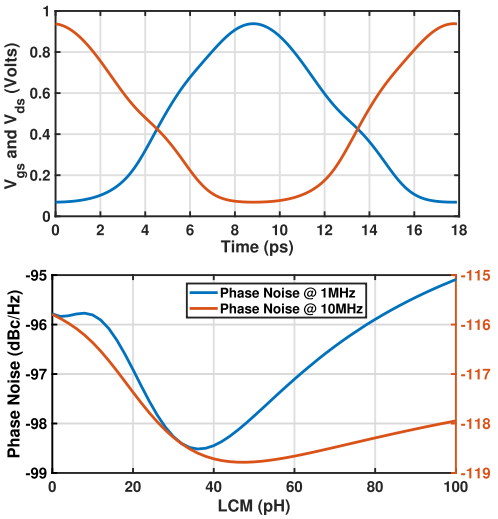


Fig. 3. (a) Transistor V_{gs} and V_{ds} and (b) phase noise variation with L_{CM} for the proposed oscillator oscillating at 56 GHz.

the secondary side for coupling and synchronizing multiple VCOs for phase noise reduction. By stacking, the peak V_{gs}, V_{ds}, and V_{gb} across each of the NMOS transistors is kept approximately below V_{DD} even while maintaining a voltage efficiency greater than 90% for the individual cross-coupled pairs [see Fig. 3(a)]. Multi-core oscillator utilizes the proposed stacked structure to accomplish reliable operation and phase noise reduction and distribute four synchronized outputs at four remote ends of the layout for LO distribution

$$f_{\text{diff}} = \frac{1}{2\pi\sqrt{(1+k)L_T(C_D+0.5C_C)}} \tag{3}$$

$$f_{\text{cm}} = \frac{1}{2\pi\sqrt{(0.5(1-k)L_T+2L_{\text{CM}})C_C}}. \tag{4}$$

To understand the functioning of the stacked oscillator, let us consider conventional exemplar NMOS oscillator working at half the nominal supply [see Fig. 2(a)] and having same Q. The differential inductance and the common-mode inductance are $L_{\text{diff}} = (1+k)L_T$ and $L_{\text{cm}} = 0.5(1-k)L_T + 2L_{\text{CM}}$, respectively. For the oscillation frequency given by (3), the peak energy stored in the tank is given by

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

$E_{\text{tank}} = 1/2C_T V_{\text{pk}}^2$. Here, V_{pk} is the peak differential amplitude. The common-mode resonance frequency is given by (4). Here, C_D and C_C are the differential- and common-mode capacitances, respectively. The common-mode capacitance is dominated by the parasitic capacitance C_{gs} of the transistor.

Stacking two of such NMOS oscillators, as shown in Fig. 2(b), stores two times the tank energy at the same resonance frequency, provided that the quality factor of the resonator is unchanged. The noise factor F remains unchanged since the transistor V_{gs} and V_{ds} waveforms across the transistors in Fig. 2(a) and (b) are the same. The assumption in Fig. 2(b) is that the oscillator is oscillating in the in-phase excitation mode. Hence, by (2), the phase noise should reduce by 3 dB compared to Fig. 2(a) based on the energy–phase–noise relation (2) in the thermal noise region [6], [13]. The power consumption doubles; hence, the FOM remains the same. For the simulations, the transformer inductance L_T, the tank capacitance C_T, the common-mode inductance L_{CM}, and the transformer coupling factor k are taken as 80 pH, 68 f, 30 pH, and 0.5, respectively. The overall tank quality factor, Q_T, is approximately 8 in both cases. It can be seen from Fig. 3(b) that approximate L_{CM} of 30 pH optimizes the phase noise performance [14], [15]. The approximate common-mode quality factor is 4. It is low due to the low-quality factor of the transistor's parasitic capacitance beyond 100-GHz frequencies.

B. Phase Noise in Stacked Oscillator

To understand why the energy–phase–noise relation (2) holds for the proposed oscillator, we provide an alternative view based on the time-domain ISF [16]–[19]. Fig. 4 shows the simulated ISF plot for the noise sources normalized to Qmax $= C_T V_{\text{pk}}$ (see also Fig. 5). It is seen that the ISFs of both the noise sources, i.e., transistor and tank loss, are reduced by a factor of 2 for the magnetically coupled stacked NMOS case compared to the reference NMOS-only oscillator. ISFs are plotted using the PXF analysis in the cadence [20]. Here, in both cases, the oscillating frequency is equal and approximately close to 56 GHz

$$L(\Delta\omega) = 10\,\text{Log}\frac{N_{L,\text{gm}} + N_{L,R_T}}{2\Delta\omega^2 V_{\text{TANK}}^2 C_T^2} \tag{5}$$

$$N_{L,(\text{gm},R_T)} = \frac{1}{T_O}\int_0^{T_O} \Gamma^2(t)\overline{i_{n,(\text{gm},R_T)}^2}dt. \tag{6}$$

Hence, the coupled resonator in the proposed stacked oscillator is two times less sensitive to phase perturbation due to noise sources compared to the LC resonator in the exemplar NMOS-only oscillator having the same Q. Since there are two times more noise sources (see Fig. 5) in the stacked case by (5) and (6), we can also intuitively see the factor 2 reduction in the phase noise, which supports the energy–phase–noise relation (2). The same result can be proved mathematically utilizing Kärtner's mathematical framework [21]–[23] for a simplistic case, as seen in Fig. 5. This hints at the equivalence between fluctuation (ISF) and the dissipation-based approach (2) [6] in the magnetically coupled oscillators.

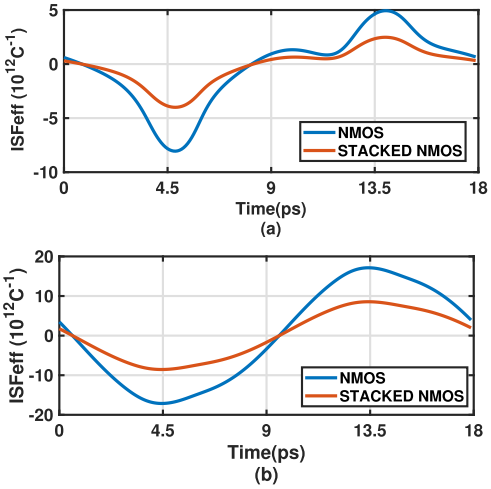


Fig. 4. Effective ISF for (a) transistor and (b) parallel tank resistance noise.

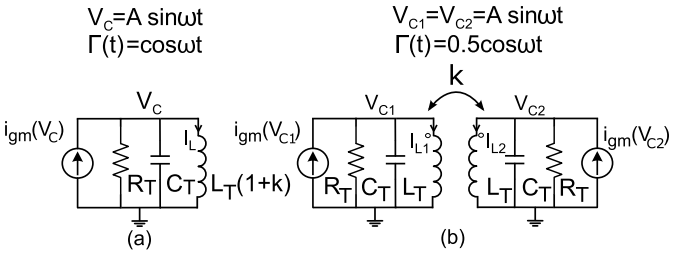


Fig. 5. (a) NMOS oscillator model. (b) Stacked NMOS oscillator model.

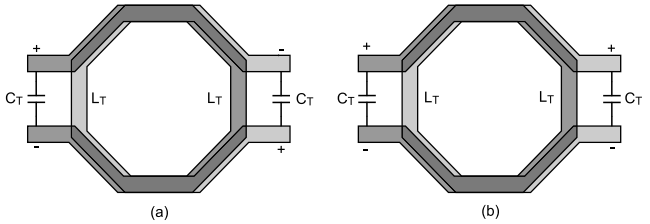


Fig. 6. Oscillator modes: (a) in-phase excitation or desired and (b) anti-phase excitation or undesired.

C. Oscillation Modes in Stacked Oscillator

$$f_{\text{mode1}} = \frac{1}{2\pi\sqrt{(1+k)L_T C_T}} \tag{7}$$

$$f_{\text{mode2}} = \frac{1}{2\pi\sqrt{(1-k)L_T C_T}}. \tag{8}$$

Fig. 6 shows the two potential modes of oscillation that can occur in the magnetic coupling of the oscillators. In the desired low-frequency mode (f_{mode1}), the effective inductance seen by each primary or secondary side increases due to positive mutual coupling, while, in the undesired mode (f_{mode2}), the negative mutual coupling causes the pole to shift to a higher frequency. The corresponding resonant modes are given by (7) and (8). A high coupling factor of close to $k \approx 0.5$ is chosen to push the pole to a much higher frequency than the desired mode. It can be seen later in Fig. 11 that the undesired pole is usually at a higher frequency than given by (8).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

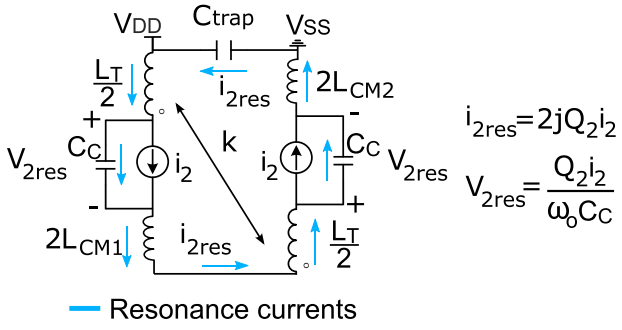


Fig. 7. Impact of the common-mode inductance mismatch.

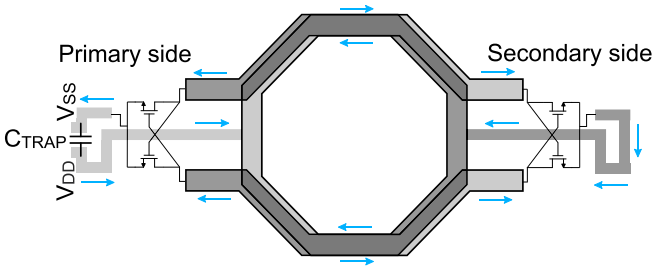


Fig. 8. Common-mode return path.

The primary reason is the increase in the coupling factor with the frequency, which pushes it to an even higher frequency. The loss of passives, especially the low-quality factor of the switched capacitors at such high frequency, makes this mode unsustainable.

D. Common-Mode Return Path and Impedance

Fig. 7 shows the common-mode currents at resonance. The common-mode inductance mismatch due to layout in the top and bottom oscillator pairs will not have an impact as long as the total common-mode inductance remains the same. Here, $V_{2\text{res}}$, $i_{2\text{res}}$, and Q_2 are second-harmonic voltage, resonance currents, and quality factor, respectively. The second-harmonic content and transistor waveforms are unaffected if the loop's common-mode inductance is held constant. The phase noise up-conversion is, thus, unaltered.

The layout of the passives used for the stacked single-core design is shown in Fig. 8. It can be seen that the return path is very well defined. Hence, the total common-mode inductance can be controlled with good accuracy, which is critical, especially for millimeter-wave oscillators. The return path of the common-mode currents is completed through a trap capacitor connected between supply and ground.

Fig. 9 shows the EM simulations of the proposed passive in the differential mode. It indicates the transformer's primary and secondary inductance, quality factor, and coupling factor, respectively. The simulated primary and secondary inductances are approximately 80 pH with a coupling factor of 0.5.

The design has no common-mode tuning because of the low-quality factor of the switched capacitors at 120 GHz. This would have reduced Z_{CM} and, thus, lower second-harmonic content in the common mode. Simulations show that the inductive common-mode impedance optimizes for phase noise across the tuning range. The performance degradation across

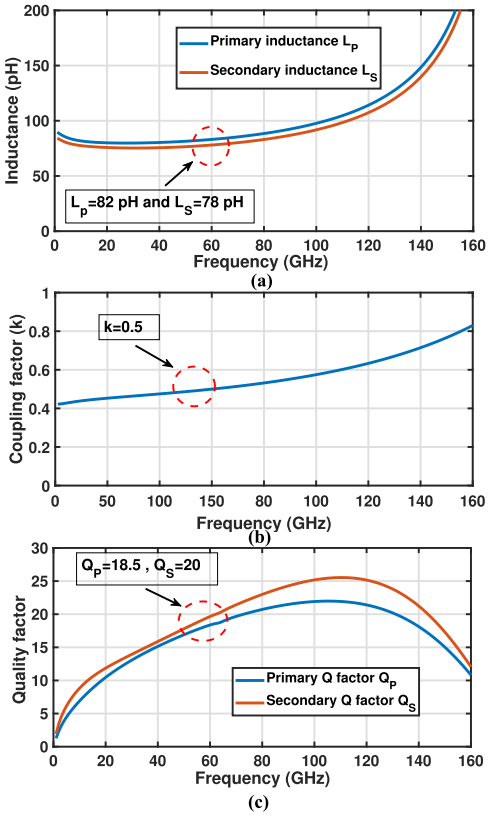


Fig. 9. (a) Transformer's primary and secondary inductances, (b) coupling factor, and (c) quality factor from the EM simulations in the differential mode.

the tuning range at 1 MHz was within 1 dB. High positive mutual coupling in the desired mode leads to the destructive coupling of fields of common-mode currents in the main inductor coil, as shown in Fig. 8. The gm cells are added at the end of the inductor stub to add an extra L_{CM} and tune the common-mode inductance to an optimum value (the stubs are needed to accommodate the capacitive tuning banks). Fig. 10 shows the common-mode inductance seen by the transistor and its quality factor. It is approximately equal to 80 pH with a quality factor of 11. Hence, the total inductance of the entire common-mode loop is controlled to an optimum value implicitly without requiring an extra-large tail inductor as required in a conventional complementary configuration.

The differential-mode impedance and the common-mode impedance after the passive layout are plotted in Fig. 11. The simulated differential- and common-mode quality factors Q_{DM} and Q_{CM} are approximately 8 and 4. The common-mode capacitance $C_C \approx 20f$ is same as the transistor C_{gs}. Transistors' parasitic capacitance and its quality factor are extracted from PSS simulations for impedance estimations. The gate resistance is the dominant factor in deciding the common mode Q_{CM}.

E. Asymmetric Loading in Stacked Oscillator

When multiple of the unit oscillators are coupled in parallel, there may be different loads on the primary and secondary sides of the transformer (see Fig. 12). Fig. 13(c) shows that the simulated phase noise is relatively constant in the

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

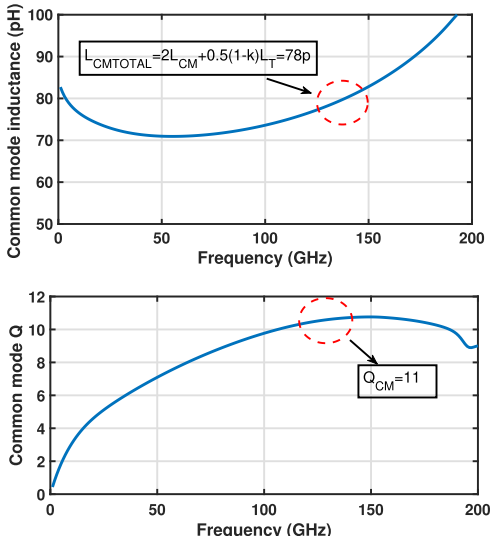


Fig. 10. (a) Total common-mode inductance and (b) its quality factor extracted from the EM simulations.

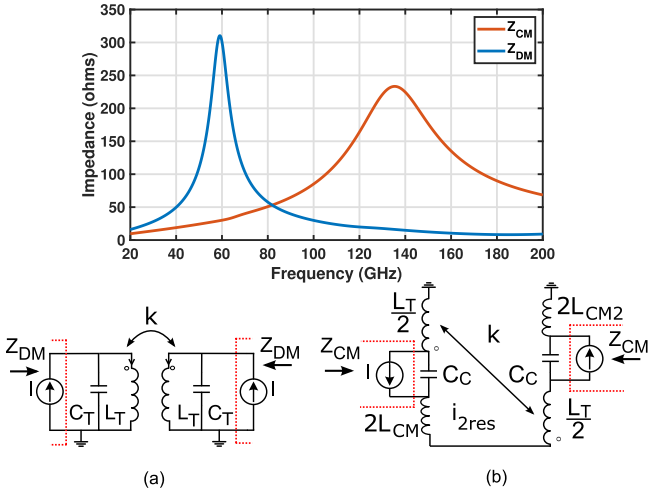


Fig. 11. (a) Differential- and (b) common-mode impedances of single core.

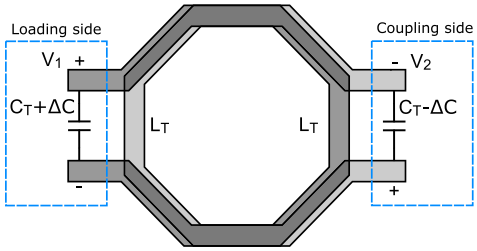


Fig. 12. Mismatch in the capacitive load resulting from asymmetric loading.

thermal region with small asymmetric loading, provided that the quality factor of the passives is unchanged. The analysis of E_{TANK} and tank Q as per (2) is done here to find the physical

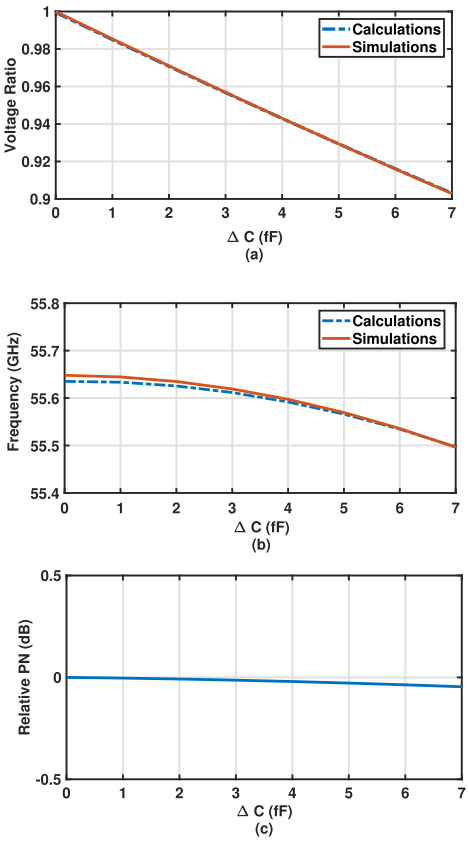


Fig. 13. (a) Simulated secondary-to-primary voltage ratio, (b) frequency, and (c) phase noise variation with mismatch capacitance.

reasoning. We here define factor γ (9)

$$\gamma = \frac{C_T - \Delta C}{C_T + \Delta C} \tag{9}$$

$$\phi_1 = L_T I_{C1} + kL_T I_{C2} = L_T I_{C1}(1 + k\alpha\gamma) \tag{10}$$

$$\phi_2 = L_T I_{C2} + kL_T I_{C1} = L_T I_{C2}\left(1 + \frac{k}{\alpha\gamma}\right) \tag{11}$$

$$L_T(C_T + \Delta C)(1 + k\alpha\gamma) = L_T(C_T - \Delta C)\left(1 + \frac{k}{\alpha\gamma}\right). \tag{12}$$

Here, $C_T = 68f$ is the overall differential tank capacitance, which includes the transistor parasitic $C_{\text{par}} = 23f$. Unbalanced loading can lead to different differential swings at the primary and secondary. Equations (10)–(12) derived from the transformer theory can be used to find the voltage ratio α and the oscillation frequency ω_o. The exact expressions of these parameters are given as follows:

$$\alpha = \frac{V_2}{V_1} = \frac{-(1-\gamma) + \sqrt{(1-\gamma)^2 + 4k^2\gamma}}{2k\gamma} \tag{13}$$

$$\omega_o = \frac{1}{\sqrt{L_T(C_T + \Delta C)(1 + k\alpha\gamma)}}. \tag{14}$$

Fig. 13(a) and (b) depicts simulated and calculated voltage ratio and frequency with the asymmetric mismatch. Fig. 13(a) indicates that each side of the transformer sees a different voltage swing. To find the reason for the unchanged phase

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

noise, we need to look into the variation of tank amplitude, $V_1 = A_1 \sin \omega_o t$ and $V_2 = A_2 \sin \omega_o t$, at loading and coupling side with mismatch capacitance ΔC

$$P_o = \frac{1}{T_O} \int_0^{T_O} 2V_o I dt \tag{15}$$

$$P_{\Delta C} = \frac{1}{T_O} \int_0^{T_O} V_1(1+\alpha) I dt. \tag{16}$$

The total average fundamental power delivered by the gm cells to the tank for the nominal case (P_O) and asymmetric loading case ($P_{\Delta C}$) is given by (15) and (16). Here, the non-linearity is assumed to be saturating; hence, for small variation in the fundamental amplitude, the fundamental current remains approximately the same. The approximate ratio between the average powers delivered by gm cells in two cases can be written as (17)

$$\frac{P_{\Delta C}}{P_o} \approx \frac{A_1(1+\alpha)}{2A_O} \tag{17}$$

$$\frac{A_1(1+\alpha)}{2A_o} = \frac{A_1^2(r_l(1+\alpha^2\gamma^2) + r_{c1} + \alpha^2\gamma^2 r_{c2})}{A_o^2(2r_c + 2r_l)}. \tag{18}$$

The average power delivered is equal to the average power dissipated in the resonant network in the form of i^2r losses. Hence, equating both of them as in (18), we can get the fundamental amplitudes A_1, $A_2 = \alpha A_1$, and the relative ratios $A_{1,2}/A_O$. Here, r_l and r_c indicate the transformer winding resistance and capacitance series resistance in the nominal case. They are approximately given by $r_l \approx (\omega_o(1+k)L_T)/(Q_L)$ and $r_c \approx 1/(\omega_o((C_T - C_{\text{par}})Q_C + C_{\text{par}}Q_{\text{par}}))$, respectively. Here, C_{par} and Q_{par} are the tank parasitic capacitance and its quality factor, which is dependent on the gate resistance. Similarly, r_{c1} and r_{c2} are the effective resistances in series with the capacitance in the asymmetric loading case at the primary and secondary sides of the transformer. They are approximated by $r_{c1} \approx 1/(\omega_o((C_T - C_{\text{par}} + \Delta C)Q_C + C_{\text{par}}Q_{\text{par}}))$ and $r_{c2} \approx 1/(\omega_o((C_T - C_{\text{par}} - \Delta C)Q_C + C_{\text{par}}Q_{\text{par}}))$.

Fig. 14 shows the variation of tank voltages and tank energy with ΔC. It can be seen that the simulated tank energy does not vary with the mismatch. Also, the power delivered by the gm cells with ΔC remains approximately the same. This can be seen intuitively in Fig. 14(a). One of the tank voltages (V_1) goes up, and the other (V_2) goes down approximately by the same magnitude. The total average power delivered by the energy restoring elements $1/T_O \int_0^{T_O}(V_1 + V_2) I dt$ approximately remains the same for small mismatch in loading. The same argument can be used to explain why energy remains the same, i.e., $E_{\text{Total}} = 1/2((C_T + \Delta C)(A_0 + \Delta A)^2 + (C_T - \Delta C)(A_0 - \Delta A)^2) \approx C_T A_0^2$. Hence, we can intuitively see that E_{tank} and $Q \propto E_{\text{tank}}/E_{\text{dissipated}}$ do not change approximately to first order. Hence, by (2), the phase noise in the thermal region is not affected by small asymmetric loading.

III. Dual-Core Oscillator

A. Parasitic Modes in the Dual-Core

Coupling two single-core VCOs (see Fig. 15) gives rise to an unwanted oscillation mode resulting from the interaction between the two single-core VCOs. This mode's dynamics

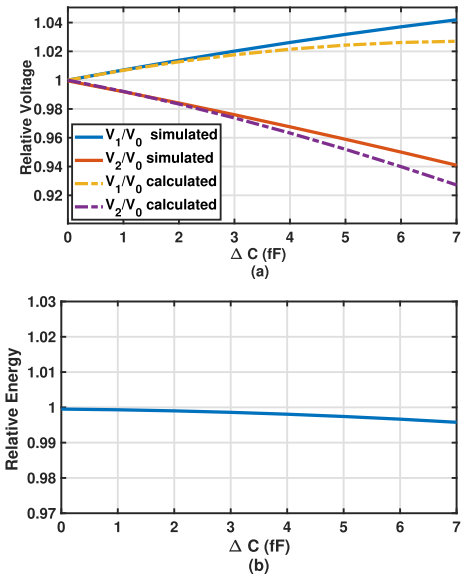


Fig. 14. (a) Relative primary $\frac{V_1}{V_o}$ and secondary $\frac{V_2}{V_o}$ voltage swings and (b) simulated relative tank energy with respect to symmetric capacitance case.

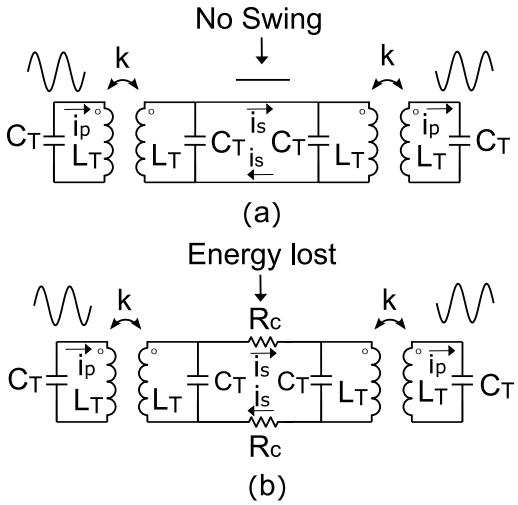


Fig. 15. (a) Parasitic mode in a dual-core and (b) its suppression.

are depicted in Fig. 15. In the desired mode, all the nodes in the coupled oscillator system oscillate in phase. In the parasitic mode3, the primary sides of each of the two single-core VCOs operate in antiphase. The secondary side of each oscillator has no swing due to the cancellation of self and mutual magnetic flux, as given by (19). Here, i_p and i_s are primary and secondary transformer currents, respectively. In this mode, the resonance currents do not flow inside the capacitors on the secondary side. The current circulates between the secondary coils on each side. Symmetric loading is assumed to simplify the analysis. The capacitors on the primary sides of each oscillator see an equivalent inductance of $L_T(1-k^2)$. The net flux in the primary winding is given by (20). The corresponding oscillation frequency is given by (21)

$$L_T i_s = M i_p \tag{19}$$

$$\phi_P = L_T i_p - M i_s = i_p L_T (1-k^2) \tag{20}$$

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

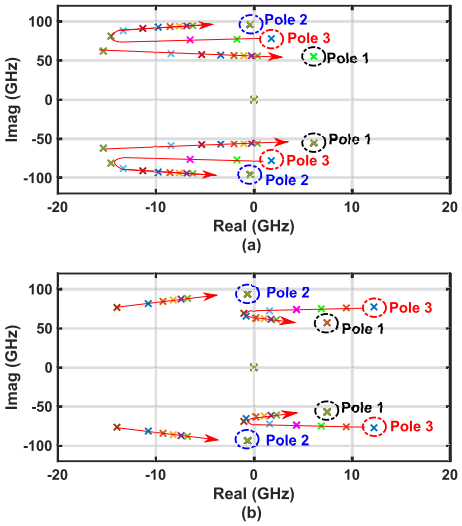


Fig. 16. Poles in (a) stacked and (b) complementary dual-core oscillators with R_C varied from 1 to 101 Ω in steps of 10 Ω.

$$f_{\text{mode3}} = \frac{1}{2\pi\sqrt{(1-k^2)L_T C_T}}. \tag{21}$$

This parasitic mode can be easily suppressed by adding a resistor between the two single-core oscillators. The current flowing through the resistor causes extra power losses in this mode, making oscillations unsustainable. The gm cells on the secondary side do not contribute to energy restoration and, thus, do not deliver any power. This is because no swing on the secondary side means no power delivered. No swing approximation is valid for low values of coupling resistance.

Fig. 16(a) shows that the third pole shifts to the left half-plane with the increase in R_C. For the pole–zero plot, we have used a frequency-independent transformer model with quality factors estimated from EM simulations of the passives. f_{mode3} given by (21) is an approximation only valid for low coupling resistances. Appendix A contains the exact differential equation explaining these other parasitic poles. As can be seen in Fig. 16(a), when coupling resistance increases, an extra fourth pole emerges. A dual-core oscillator with finite coupling resistance is an eighth-order network. When the coupling resistance is low, there are three pairs of complex conjugate poles and two real poles, making a total of eight. However, as the value of coupling resistance increases, we get four pairs of complex conjugate poles.

One of these parasitic poles (poles 3 and 4) approaches the first desired pole, and the other one approaches the second undesired pole as coupling resistance increases. This is intuitive since one can imagine that, when the coupling resistance tends to be large, there is limited interaction current flowing between the oscillators. The poles of the oscillator approach those of a single-core oscillator. Here, for resistances larger than 100 Ω, one of the poles approaches the right half-plane. Hence, the design value should be chosen somewhere between 10 and 100 Ω. In the present design, a resistance of 10 Ω is used to enable strong coupling between the oscillators and, thus, less mismatch sensitivity to the phase noise. Since the gm cells on the secondary side of the transformer do not contribute due to no swing, the value of the coupling resistor needed to suppress parasitic mode is low.

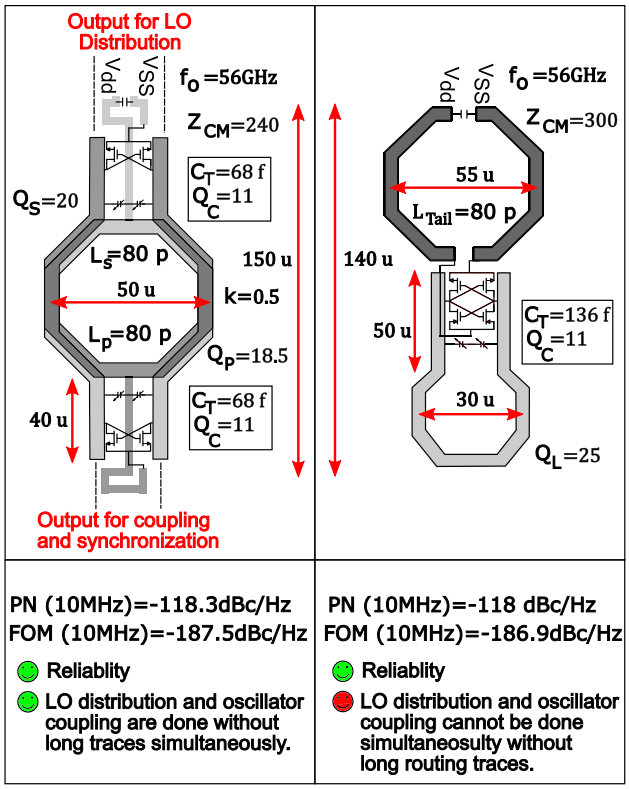


Fig. 17. Comparison between the proposed single-core and conventional single-core complementary oscillators.

B. Comparison With Conventional Complementary

Fig. 17 compares our proposed single-core with the conventional complementary oscillator. For first-order comparison, both the oscillators are designed for approximately the same phase noise at 56 GHz. Hence, the complementary oscillator is realized with an inductance of $0.5L_T(1+k) = 60$ pH. The actual phase–noise would be slightly less in the conventional complementary since an optimistic stub size is used. For a fair comparison, the stub size should be double that of a single coil in the proposed oscillator. However, it was not possible to realize such a small value of the inductor with increased stub size. The transistor widths used for the comparison are the same. It is seen that the phase noise performance is approximately the same as the proposed oscillator. A slightly better FOM in the proposed oscillator can be attributed to the higher effective primary and secondary inductance quality factors, which is $(\omega(1+k)L_{p,s})/(r_{s,p}) = (1+k)Q_{p,s}$ than the 60-pH inductor and slightly lower f_t and $f_{\max}$ of the PMOS compared to NMOS.

Our proposed single-core structure helps us to simultaneously couple multiple oscillators and efficiently distribute the LO to remote ends without long routing traces and additional buffering. This feature in the multi-core version can help with power-efficient LO distribution to TX and RX, which are

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

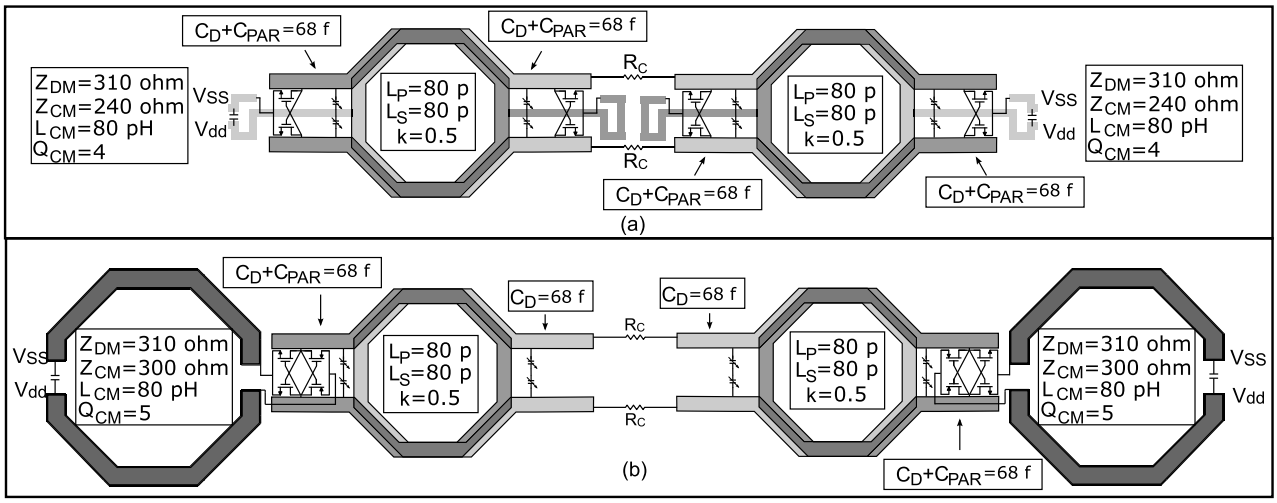


Fig. 18. (a) Proposed dual-core based on the proposed stacked oscillator. (b) Alternative dual-core based on the conventional complementary oscillator.

usually hundreds of microns apart. A very long route trace to the LO buffer will increase LO loading and reduce tank Q (series L with C). The parasitic capacitances will also increase with a very long coupling trace. Simulations also show that large coupling inductance results in increased flicker noise up-conversion in the presence of a mismatch.

Fig. 18 compares dual-core based on the proposed topology [see Fig. 18(a)] with dual-core based on a complementary PMOS–NMOS core [see Fig. 18(b)]. The complementary oscillator is designed as the proposed oscillator to enable simultaneous LO distribution on the primary and oscillator coupling on the secondary. C_T, L_T, and k are kept the same as in the proposed circuit for the analysis of the same phase noise performance. As can be seen from Fig. 18(b), the conventional complementary oscillator will require an extra tail inductor, approximately the same as the total common-mode inductance of the proposed stacked oscillator, i.e., $0.5L_T+2L_{\mathrm{CM}} \approx 80$ pH. The tail inductor, in this case, results in an increased area without storing any significant tank energy (less than 1% of E_{Tank}). With an 80-pH tail inductor having a 55 μm diameter, which resonates with the parasitics, we estimate a 50% increase in the multi-core area. The multi-core area is estimated to increase from 0.35×0.35 mm^2 to 0.43×0.43 mm^2. This area also includes that of decoupling ring. Besides that, it impedes distributing the LO on the primary side of the oscillator. This is because the distance to the output buffer increases since this inductor is present between the oscillator and the LO buffer.

The coupling resistor needed to suppress the parasitic mode is low in value in the proposed dual-core because the gm cells on the secondary side do not contribute to any energy restoration. However, now, when we look at the implementation using conventional complementary, we can see that all the energy-restoring gm cells are concentrated on the primary side of the transformer. Hence, intuitively, we can conclude that a higher value of coupling resistor is required for mode suppression. This is also seen in Fig. 16(b) that coupling resistance between 50 and 60 Ω is required to suppress the parasitic mode3. It makes the oscillator more sensitive to tank

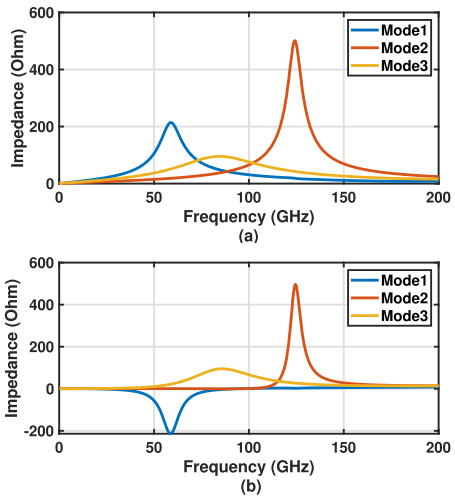


Fig. 19. (a) Magnitude and (b) real parts of the tank impedance.

mismatches. Besides this, the mode suppression works only for narrow values of resistance, somewhere between 50 and 60 Ω [see Fig. 16(b)]. There is a danger of putting the poles to the right half-plane due to process variations, which may result due to variation in gm and the coupling resistance R_C.

Fig. 19 shows the magnitude and real part of the effective tank impedance after EM extraction of the layout of the passives. The second pole has shifted to a higher frequency because of the increase in coupling factor with frequency. Only the desired mode1 has a negative real impedance.

C. Impact of Mismatch in Dual Core

Fig. 20 shows the schematic used to study the impact of loading mismatch on the phase noise performance in the dual-core, and Fig. 21 shows the phase noise penalty at 1- and 10-MHz offsets from the carrier. It can be seen that 10–20 Ω of coupling impedance results in negligible impact. Any mismatch, intuitively, will result in an interaction current

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

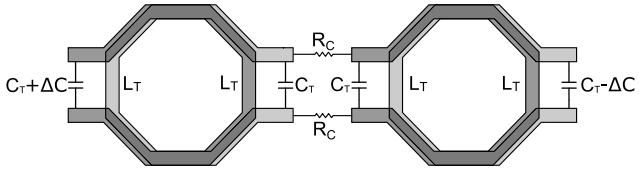


Fig. 20. Schematic for the mismatch analysis of the dual-core oscillator.

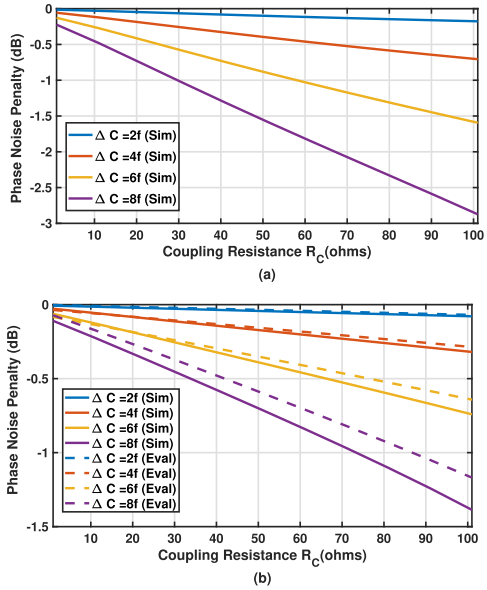


Fig. 21. Phase noise penalty at (a) 1-MHz offset simulated and (b) 10-MHz offset simulated and evaluated from state vectors in the simulations.

between the oscillators. The higher the coupling resistance, the more power will be lost. Hence, intuitively, we can expect a drop in both the tank energy and quality factor.

In Section II-B, we saw the equivalence between the fluctuation (ISF) and dissipation-based methods [see (2)] for transformer-coupled oscillators. An ISF-based analysis [1] was done to analyze phase noise degradation in the dual-core coupled oscillator resonator. After mathematical treatment using Kärtner's mathematical framework, it was derived that phase noise is inversely proportional to a^3. Here, "a" is the ratio of the tank amplitude to that of a nominal case in the absence of mismatch. The same result can be seen using the dissipation-based method, utilizing (2). Factor a^2 comes from the reduction in tank energy, and the other factor a comes from the reduction in the effective quality factor. The effective quality factor of the entire coupled oscillator system is reduced because of the current flowing through the coupling resistor in the presence of a tank mismatch. This also hints at an underlying unity between the fluctuation-based methods and dissipation-based methods for coupled oscillators in general.

Fig. 21(b) shows both simulated and evaluated phase noise degradation at a 10-MHz offset. We assume that, here, at this offset, the noise is dominated by thermal noise. To first prove that (2) is valid, we extract state vectors, i.e., currents, voltages, and the coupling current of the dual-core coupled resonator

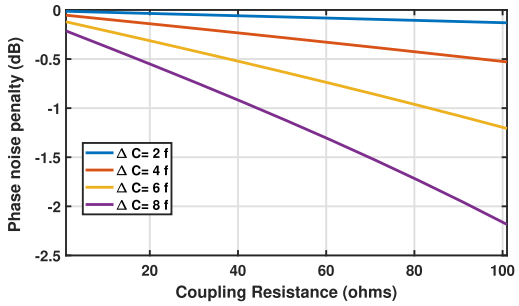


Fig. 22. Phase noise penalty (thermal) predicted from the mathematical model.

in the presence of mismatch from the simulation. The total energy stored (22) can be evaluated by summing the average energy stored in all the energy storing elements (capacitor, inductance, and mutual inductance). The total dissipated power [see (23)] is calculated by summing the power dissipated in the loss elements using the state vectors derived from simulations. From these two equations, we can estimate the $E_{\text{TANK}}Q$ product and, thus, relative degradation by (2). Since the frequency and, hence, the common-mode impedance do not change, we assume that the noise factor does not change to first order. Fig. 21(b) proves our hypothesis since the simulated relative degradation in the phase noise almost exactly matches the evaluated degradation. The small difference is due to traces of flicker noise present at a 10-MHz offset. Besides, in the simulations, higher flicker noise up-conversion was also seen when the coupling impedance increases in the presence of mismatch

$$E_{\text{Total}} = \sum \bar{E}_{L,M,C} \tag{22}$$

$$P_{\text{diss}} = \sum P_{L,C,R_C}. \tag{23}$$

Fig. 22 shows phase noise degradation predicted from the approximate mathematical model given in Appendix B. Here, the state vectors are derived using transformer theory and power-balancing equations, as done in Section II-E. The difference from the simulated value comes from the approximation of saturating non-linearity, i.e., current driven by gm cells does not change with amplitude by the first order. The fundamental physics behind the phase noise degradation in the quad-core is the same as that of the dual-core. Current will flow from oscillators showing slightly inductive impedances to those showing slightly capacitive impedance at the frequency of resonance. All these currents will flow through the coupling resistors, decreasing tank energy and Q of the overall coupled oscillator system. The physics is very similar in that we have to treat a coupled oscillator as a single oscillator that has a certain energy and quality factor.

IV. Quad-Core Implementation

In the quad-core VCO (see Fig. 23), four of the single-core VCOs are coupled on the secondary side of the transformer, and the primary side is used for LO distribution. The possible oscillation modes are the same as those of the dual-core. For coarse tuning, a 4-bit capacitor bank with an approximately

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

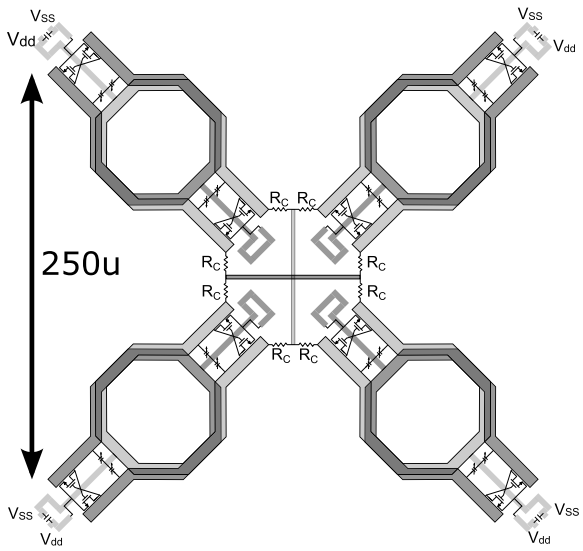


Fig. 23. Quad-core VCO layout and implementation after rotation.

total capacitance of 45 f is used on each primary and secondary. For fine-tuning, a small AMOS varactor is placed on the primary side. This was done so as not to exceed the reliability limits of the varactor, i.e., $|V_{\rm gb}| < 0.9$ V. Transformers are implemented on high-conductivity top-copper (M9) and aluminum layers (AP). The approximate transformer primary and secondary quality factor is $Q_L \approx 20$. The coupling resistors are implemented in a higher resistivity layer (M7). Cross-coupled gm cells on the secondary of the transformer in each stacked core are implemented with transistors in deep-n-well with substrate connected to the source for reliable operation, i.e., $|V_{\rm gb}| < 0.9$ V. One of the four oscillator cores is loaded on the primary side by a two-stage LO buffer (for test purposes), while the other three outputs are loaded with dummy loads. The LO buffer was designed to drive approximately 2–3 dBm of output power on a 50-Ω load for frequency and phase noise measurements.

Since the transistors in lower sub-micrometer technologies cannot be rotated (their polysilicon gates are allowed in the vertical direction only), each of the unit oscillators is made at an angle $90(n-1)+45°$ to the vertical axis to maintain symmetry. The separation between horizontal, vertical, and diagonal outputs is 250, 250, and 350 μm. Each of the outputs can be used to distribute LO to TX and RX. A supply-decoupling ring, which surrounds the oscillator, is used to distribute supply and ground. The oscillator core area is about 0.12 mm^2.

V. Measurement Results

Fig. 24 shows the micrograph of the fabricated chip in 28-nm CMOS. The prototype oscillator consumes about 12–15.3 mW across the frequency tuning range while working at a nominal supply of 0.9 V. The working frequency range (see Fig. 25) of the oscillator is 55.96–62.83 GHz, which corresponds to a tuning range of 13.4%. The spectrum was measured using the Rohde&Schwarz FSU67 spectrum analyzer. Fig. 26 shows the output power over the tuning range.

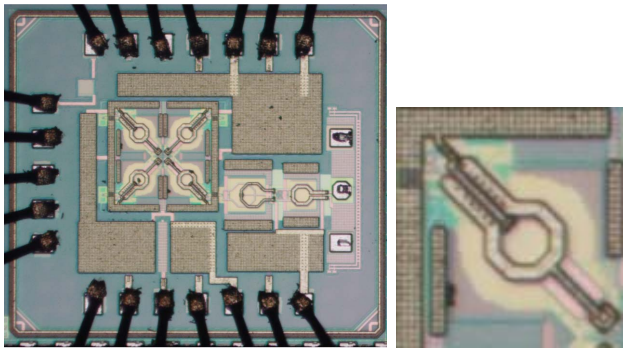


Fig. 24. Micrograph of the prototype fabricated in 28-nm CMOS.

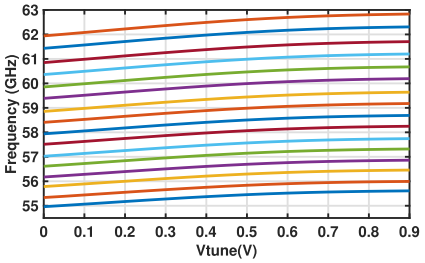


Fig. 25. Measured oscillator tuning curves of the prototype oscillator.

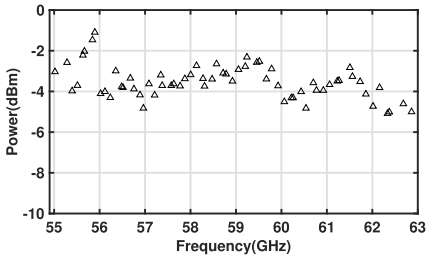


Fig. 26. Output power measured by the spectrum analyzer after cable losses.

This output power is measured after the RF cables, which have approximately 5–6 dB of power loss, and such a power level is good for accurate measurements of phase noise.

Fig. 27 shows the phase noise at 55 GHz. The phase noise is measured utilizing an AGT 5052B signal analyzer. The downconverted IF output from the spectrum analyzer is fed to the signal analyzer. The phase noise measured at 1- and 10-MHz offsets from the carrier is −95 and −121.7 dBc/Hz, respectively, corresponding to FOMs of 178 and 184.7 dBc/Hz at 1 and 10 MHz. The observed flicker noise corner is close to 5 MHz, indicating that, at 1 MHz, the noise is dominated by flicker noise up-conversion and at 10 MHz dominated by thermal noise.

Fig. 28 shows the phase noise variation across the tuning range at 1- and 10-MHz offsets from the carrier. The phase noise at 1- and 10-MHz offsets vary from −91.3 to −95 and −118.5 to −122 dBc/Hz approximately, respectively. The corresponding variations in FOM (see Fig. 29) at 1 and 10 MHz are 176.5–178 and 183.5–185 dBc/Hz, respectively. The phase noise plots are taken across the tuning range for

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

TABLE I

PERFORMANCE SUMMARY OF STATE-OF-THE-ART OSCILLATORS

Reference	This work	ESSCIRC'17 [24]	JSSC'18 [4]	TMTT'16 [25]	JSSC'19 [3]	ISSCC'2020 [2]	ISSCC'21 [12]
Tech	28nm CMOS	45nm SOI	28nm CMOS	65nm CMOS	28nm CMOS	65nm CMOS	65nm CMOS
Number of Cores	4	2[c]	1[c]	2[b]	1[c]	1[c]	4
Architecture	Stacked NMOS	NMOS	NMOS	NMOS	NMOS	NMOS	NMOS
Tuning Range (GHz)	54.96-62.83 13.4%	56-67 18%	27.3-31.2 14%	56-60.4 9.1%	57.5-67.2 15.6%	54.9-63.5 14.5%	52.4-60.4 14.2%
Supply Voltage(V)	0.9	1	1	1	0.9	0.9	0.55
Power (mW)	15.3	90	22	24	29.5	9	23.6
PN (dBc/Hz) @ 1MHz	-95	-97.7[a]	-100[a]	-95	-99.6[a]	-101.4 [a]	-105.3[a]
PN (dBc/Hz) @ 10MHz	-121.7	-121[a]	-120[a]	-119	-119.1[a]	-119.9[a]	-130.6[a]
FOM (dBc/Hz) @ 1MHz	178	173	181	176	179.8	186.7	186.5
FOM (dBc/Hz)@ 10MHz	184.7	176	184	179.8	179.3	185.2	192

[a]Referred to 55GHz, [b]QVCO, [c]Non-fundamental VCOs based on harmonic extraction, $FOM = -PN + 20log_{10}(f_o/\Delta f) - 10log_{10}(P_{DC}/1mW)$

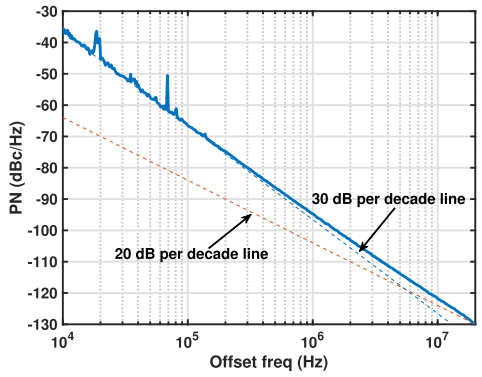


Fig. 27. Phase noise plot of the prototype measured at 55.23 GHz.

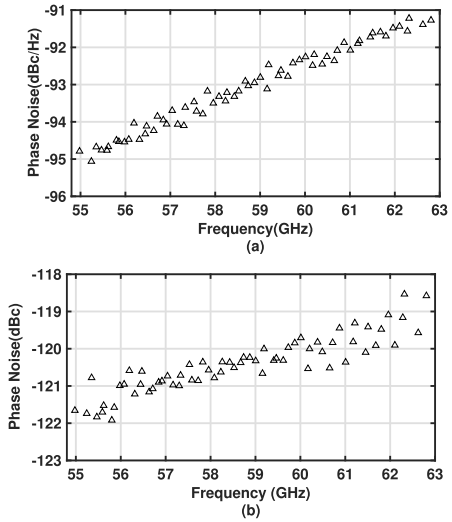


Fig. 28. Phase noise at (a) 1 and (b) 10 MHz across the tuning range.

different tuning settings of the capacitor bank and the analog varactor.

The flicker noise up-conversion was significantly higher than predicted from simulations. For example, at 1 MHz, the difference was approximately 7–8 dB at 55 GHz. At 10-MHz

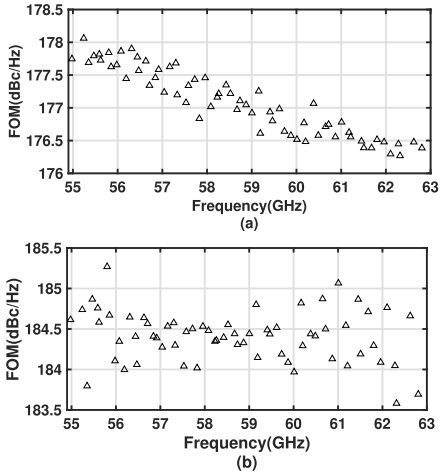


Fig. 29. Measured FOM at (a) 1- and (b) 10-MHz offsets from the carrier.

offsets, the simulations and measurements were accurate to within 1 dB. An intentional mismatch was created between the different cores to analyze the issue. Phase noise degradation was seen at 10-MHz offset but not at lower offsets of less than 1 MHz. This contradicted simulations, which showed intentional mismatch, resulted in higher flicker noise up-conversion. Hence, it is suspected that AM–PM conversion due to varactors and transistors, which were not captured in the simulations, might be the issue. The phase–noise performance did not change with two different power supplies, indicating that the phase–noise is coming from the DUT.

As can be seen in Table I, performance compares well with the state of the art while working at the nominal supply voltage of the technology. Furthermore, this work also takes care of the reliability of the devices (stack of NMOS under core voltage), which is a concern when we scale down to lower sub-micrometer technologies. Compared to [2] and [4], which work at nominal supply voltage, this work has similar performance at a 10-MHz offset. At a 1-MHz offset, our performance degrades due to flicker noise up-conversion. However, [2] and [4] are not fundamental frequency oscillators. Device reliability will also be an issue in [2] and [4] due to nominal

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

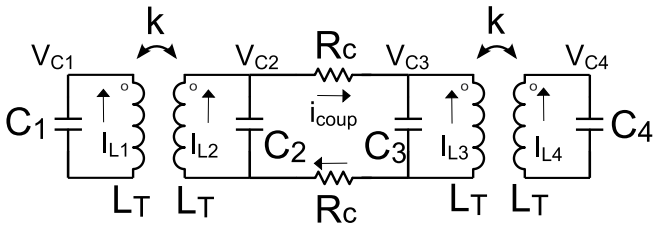


Fig. 30. Oscillator state variables and resonance voltage and currents.

supply operation with thin oxide devices. Jia *et al.* [12] show better performance than us. The possible reasons could be increased Q due to circular geometry, passive amplification, and a degrading noise factor at lower technologies. However, it is operated at 0.55 V, which is less than twice the nominal supply. Hence, a degradation in FOM is expected in a practical scenario. This is due to the high dropout voltage in regulators needed to supply it. As per the author's knowledge, this work is the first fundamental multi-core VCO at such high frequencies, which takes care of device reliability, and also the first fundamental multi-core VCO at 60 GHz in 28-nm CMOS.

VI. Conclusion

This work demonstrates a 55–63-GHz fundamental multicore-VCO in 28-nm CMOS utilizing stacked NMOS resonators for phase noise reduction and reliable operation under nominal supply. Four core oscillators, each stacking two NMOS-only LC-tank stages, are coupled in a configuration that simplifies LO distribution. The common-mode harmonic impedance is controlled implicitly without any extra-large tail filtering inductor. The alternative implementation using a conventional complementary oscillator used to simplify LO distribution needed an extra tail filtering inductor. Such a tail inductor, required for a conventional complementary core, not only increases silicon area but also acts as an impediment to the LO distribution. It was also seen that the distributed energy restoring elements in the stacked oscillator also result in robust parasitic mode suppression compared to the alternative implementation based on the complementary oscillator. Phase noise and FOM are comparable to recently published oscillators working at nominal supply, despite operating at the fundamental frequency, while stacking NMOS stages in each oscillator core guarantees reliable operation under the nominal supply.

Appendix A

In this appendix, the differential equation corresponding to the stray modes is derived. The transformer theory is used to derive the exact frequencies of the stray modes (see Fig. 30). The primary and secondary voltages of each oscillator are related by (A1) and (A2). For simplification, all the capacitances are taken as C. Similarly, transformer secondary currents and voltages are related by (A3) and (A4)

$$V_{C1} = \frac{kV_{C2}}{1 + s^2\mathrm{LC}_1(1-k^2)} \tag{A1}$$

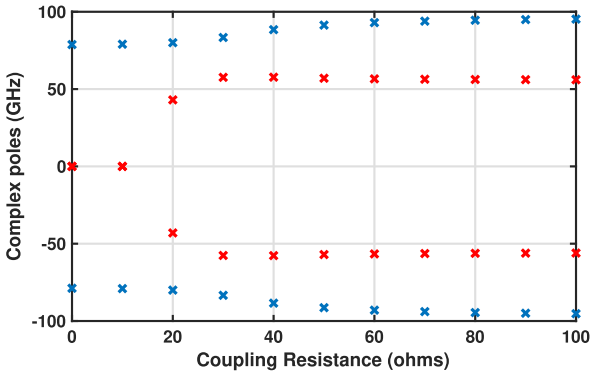


Fig. 31. Stray complex conjugate poles with coupling resistance R_C.

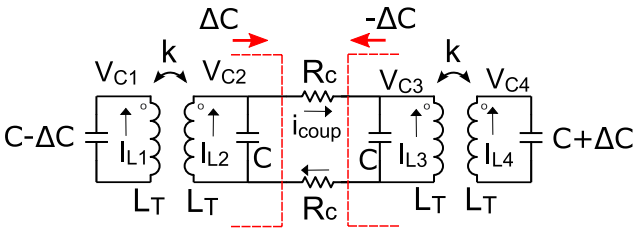


Fig. 32. Capacitive load seen on the coupling side with mismatch.

$$V_{C4} = \frac{kV_{C3}}{1 + s^2\mathrm{LC}_4(1-k^2)} \tag{A2}$$

$$I_{L2} = \frac{-V_{C2}(1 + s^2\mathrm{LC}_1)}{sL(1 + s^2\mathrm{LC}_1(1-k^2))} \tag{A3}$$

$$I_{L3} = \frac{-V_{C3}(1 + s^2\mathrm{LC}_4)}{sL(1 + s^2\mathrm{LC}_4(1-k^2))}. \tag{A4}$$

The coupling current can be related to transformer secondary currents by (A5). Utilizing (A3)–(A5), the final differential equation describing the dynamics of the oscillator can be derived (A6)

$$\frac{(I_{L2} - I_{L3})}{2(1 + sCR_C)} = \frac{V_{C2} - V_{C3}}{2R_C} \tag{A5}$$

$$(V_{C2} - V_{C3})\left(s^4L^2C^2R_C(1-k^2)\right.$$

$$\left. + s^3L^2C(1-k^2) + 2s^2\mathrm{LCR}_C + sL + R_C\right) = 0. \tag{A6}$$

When both oscillators oscillate in a mode other than either mode1 or mode2, then $V_{C2} \neq V_{C3}$. The quartic equation in (A6) can be used to find other possible modes. The complex conjugate roots or modes, which are the solution to (A6) due to oscillator interaction, are plotted in Fig. 31.

Appendix B

This appendix discusses an approximate method for quantitative estimation of phase noise degradation in the thermal region in the presence of capacitor mismatch (see Fig. 32). In the presence of a mismatch, the currents flowing through the coupling resistor result in the reduction of tank energy and the overall quality factor of the coupled oscillator system.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

We use the energy–phase-noise relation (2) to estimate phase noise degradation. It can be proved from the transformer theory that the approximate oscillation frequency is given by (A7) in the presence of a small loading mismatch. The coupling current i_{coup} has components in-phase (i_r) and 90° out-of-phase (i_c) with voltage V_{C2}. Capacitive coupling current flowing out-of-phase with V_{C2} flowing through the coupling resistor is approximately given by $i_c = sV_{C2}\Delta C$. This is intuitive because each of the oscillators sees equal and opposite capacitive loading (ΔC) on the coupling side. For the in-phase component current i_r, we have to use power-balancing equations (A8) and (A9)

$$\omega_o \approx \frac{1}{\sqrt{(\text{LC}(1+k)}} \tag{A7}$$

$$0.5(-V_{C2}i_r + (V_{C2} + V_{C1})I) = P_{\text{osc1}} \tag{A8}$$

$$0.5(V_{C2}i_r + (V_{C3} + V_{C4})I) = P_{\text{osc2}} + 0.5i_c^2(2R_c) \tag{A9}$$

$$V_{C2} \approx V_{C3} + 2i_r R_C \tag{A10}$$

where P_{osc1} and P_{osc2} are the average powers dissipated in the two oscillators. I is the drive current of the energy restoring gm cells, which can be estimated based on the Q and the tank voltage in the absence of mismatch. Using (A1), (A2), and the power-balancing equations, one can estimate the voltages and currents in the coupled oscillator network. The approximation in this analysis is that the drive approximately remains constant with small changes in the amplitude. Once the state variables are calculated, one can determine the resonator energy, which is the sum of the average energies of all the energy-storing elements. The power loss can be calculated by calculating the power lost in all the passive elements and the coupling resistor.

ACKNOWLEDGMENT

The authors would like to thank Z. Zhong, S. V. Winckel, A. Dewilde, H. Suys, and L. Pauwels (imec) for measurement support and fruitful discussions. They would also like to thank Sarthak Sharma (University of Pavia) for general discussions on oscillators.

REFERENCES

[1] L. Iotti, A. Mazzanti, and F. Svelto, "Insights into phase-noise scaling in switch-coupled multi-core LC VCOs for E-band adaptive modulation links," *IEEE J. Solid-State Circuits*, vol. 52, no. 7, pp. 1703–1718, Jul. 2017.

[2] C. Fan, J. Yin, C.-C. Lim, P.-I. Mak, and R. P. Martins, "A 9 mW 54.9-to-63.5 GHz current-reuse LO generator with a 186.7 dBc/Hz FoM by unifying a 20 GHz 3rd-harmonic-rich current-output VCO, a harmonic-current filter and a 60 GHz TIA," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2020, pp. 282–284.

[3] Z. Zong, P. Chen, and R. B. Staszewski, "A low-noise fractional-N digital frequency synthesizer with implicit frequency tripling for mm-wave applications," *IEEE J. Solid-State Circuits*, vol. 54, no. 3, pp. 755–767, Mar. 2019.

[4] Y. Hu, T. Siriburanon, and R. B. Staszewski, "A low-flicker-noise 30-GHz class-F23 oscillator in 28-nm CMOS using implicit resonance and explicit common-mode return path," *IEEE J. Solid-State Circuits*, vol. 53, no. 7, pp. 1977–1987, Jul. 2018.

[5] S. Balamurali, G. Mangraviti, C.-H. Tsai, P. Wambacq, and J. Craninckx, "A 55–63 GHz fundamental quad-core VCO with NMOS-only stacked oscillator in 28 nm CMOS," in *Proc. ESSCIRC*, 2021, pp. 295–298.

[6] D. Ham and A. Hajimiri, "Virtual damping and Einstein relation in oscillators," *IEEE J. Solid-State Circuits*, vol. 38, no. 3, pp. 407–418, Mar. 2003.

[7] D. Murphy and H. Darabi, "A 27-GHz quad-core CMOS oscillator with no mode ambiguity," *IEEE J. Solid-State Circuits*, vol. 53, no. 11, pp. 3208–3216, Nov. 2018.

[8] F. Padovan, F. Quadrelli, M. Bassi, M. Tiebout, and A. Bevilacqua, "A quad-core 15 GHz BiCMOS VCO with −124 dBc/Hz phase noise at 1 MHz offset, −189 dBc/Hz FOM, and robust to multimode concurrent oscillations," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2018, pp. 376–378.

[9] S. A.-R. Ahmadi-Mehr, M. Tohidian, and R. B. Staszewski, "Analysis and design of a multi-core oscillator for ultra-low phase noise," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 63, no. 4, pp. 529–539, Apr. 2016.

[10] L. Tomasin, G. Boi, F. Padovan, and A. Bevilacqua, "A 10.7–14.1 GHz reconfigurable octacore DCO with −126 dBc/Hz phase noise at 1 MHz offset in 28 nm CMOS," in *Proc. IEEE Radio Freq. Integr. Circuits Symp.*, Jun. 2021, pp. 179–182.

[11] A. Moroni, R. Genesi, and D. Manstretta, "Analysis and design of a 54 GHz distributed 'hybrid' wave oscillator array with quadrature outputs," *IEEE J. Solid-State Circuits*, vol. 49, no. 5, pp. 1158–1172, May 2014.

[12] H. Jia, W. Deng, P. Guan, Z. Wang, and B. Chi, "A 60 GHz 186.5 dBc/Hz FoM quad-core fundamental VCO using circular triple-coupled transformer with no mode ambiguity in 65 nm CMOS," in *IEEE Int. Solid-State Circuits Conf. (ISSCC) Dig. Tech. Papers*, Feb. 2021, pp. 1–3.

[13] A. Mazzanti and A. Bevilacqua, "On the phase noise performance of transformer-based CMOS differential-pair harmonic oscillators," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 62, no. 9, pp. 2334–2341, Sep. 2015.

[14] D. Murphy, H. Darabi, and H. Wu, "Implicit common-mode resonance in LC oscillators," *IEEE J. Solid-State Circuits*, vol. 52, no. 3, pp. 812–821, Mar. 2017.

[15] M. Babaie and R. B. Staszewski, "An ultra-low phase noise class-F 2 CMOS oscillator with 191 dBc/Hz FoM and long-term reliability," *IEEE J. Solid-State Circuits*, vol. 50, no. 3, pp. 679–692, Mar. 2015.

[16] J. Bank, "A harmonic-oscillator design methodology based on describing functions," Ph.D. dissertation, Dept. Signals Syst., School Elect. Eng., Chalmers Univ. Techn., Chalmers, Sweden, 2006.

[17] D. Murphy, J. J. Rael, and A. A. Abidi, "Phase noise in LC oscillators: A phasor-based analysis of a general result and of loaded Q," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 57, no. 6, pp. 1187–1203, Jun. 2010.

[18] A. Hajimiri and T. H. Lee, "A general theory of phase noise in electrical oscillators," *IEEE J. Solid-State Circuits*, vol. 33, no. 2, pp. 179–194, Feb. 1998.

[19] A. Mazzanti and P. Andreani, "Class-C harmonic CMOS VCOs, with a general result on phase noise," *IEEE J. Solid-State Circuits*, vol. 43, no. 12, pp. 2716–2729, Dec. 2008.

[20] Y. Hu, T. Siriburanon, and R. B. Staszewski, "Intuitive understanding of flicker noise reduction via narrowing of conduction angle in voltage-biased oscillators," in *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 66, no. 12, pp. 1962–1966, Dec. 2019.

[21] F. X. Kaertner, "Determination of the correlation spectrum of oscillators with low noise," *IEEE Trans. Microw. Theory Techn.*, vol. 37, no. 1, pp. 90–101, Jan. 1989.

[22] F. X. Kärtner, "Analysis of white and $f^{-\alpha}$ noise in oscillators," *Int. J. Circuit Theory Appl.*, vol. 18, no. 5, pp. 485–519, Sep./Oct. 1990.

[23] P. Andreani, "A time-variant analysis of the $1/f^2$ phase noise in CMOS parallel LC-tank quadrature oscillators," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 53, no. 8, pp. 1749–1760, Aug. 2006.

[24] J. Rimmelspacher, R. Weigel, A. Hagelauer, and V. Issakov, "A quad-core 60 GHz push-push 45 nm SOI CMOS VCO with −101.7 dBc/Hz phase noise at 1 MHz offset, 19% continuous FTR and −187dBc/Hz FoMT," in *Proc. ESSCIRC*, 2018, pp. 138–141.

[25] T. Xi, S. Guo, P. Gui, D. Huang, Y. Fan, and M. Morgan, "Low-phase-noise 54-GHz transformer-coupled quadrature VCO and 76-/90-GHz VCOs in 65-nm CMOS," *IEEE Trans. Microw. Theory Techn.*, vol. 64, no. 7, pp. 2091–2103, Jul. 2016.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.



Sriram Balamurali (Student Member, IEEE) received the B.E. degree in electronics and communications engineering from the Delhi College of Engineering, New Delhi, India, in 2012, and the M.Sc. degree *(cum laude)* from the Delft University of Technology, Delft, The Netherlands, in 2018. He is currently pursuing the Ph.D. degree in RF/mm-wave frequency synthesis with Vrije Universiteit Brussel, Brussels, Belgium, in collaboration with imec, Leuven, Belgium.

He was a Mixed-Signal Engineer with Freescale/NXP Semiconductor, Noida, India, from 2012 to 2016. From 2017 to 2018, he has worked on cryogenic low-noise amplifiers (LNAs) targeted for quantum computing applications. His current research interests include RF/mm-wave frequency synthesis and transceivers. His interests also lie in exploring and understanding the physics of noise in electronic oscillators and amplifiers.



Giovanni Mangraviti (Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical engineering from the University of Messina, Messina, Italy, in 2005 and 2008, respectively, and the Ph.D. degree from the Free University of Brussels (Vrije Universiteit Brussel), Brussels, Belgium, in 2015.

In 2007, he was an Intern with imec, Leuven, Belgium, where he was involved with the characterization and investigation of advanced low-*k* dielectric materials for back-end-of-line. In 2009, he joined the mm-Wave R&D Team, imec. Investigation, design, measurement, and advance of mm-wave integrated core circuits and transceivers currently involve his research activities and passion.

Dr. Mangraviti has been serving as a TPC Member of the IEEE Asian Solid-State Circuits Conference since 2019.



Cheng-Hsueh Tsai (Student Member, IEEE) received the B.S. and M.S. degrees in electrical engineering from National Taiwan University (NTU), Taipei, Taiwan, in 2011 and 2013, respectively. He is currently pursuing the Ph.D. degree with the Department of Electronics and Informatics (ETRO), Vrije Universiteit Brussel (VUB), Brussels, Belgium, in collaboration with imec, Leuven, Belgium.

In 2020, he joined Nordic Semiconductor, Oslo, Norway, where he is working in analog front-end design for Bluetooth. His research interests include data converters, frequency synthesis, and transceivers for mm-wave wireless communication.



Piet Wambacq (Senior Member, IEEE) received the Ph.D. degree in electrical engineering from the Katholieke Universiteit Leuven, Leuven, Belgium, in 1996.

Since 1996, he has been with imec, Leuven, where he is currently a fellow. He is also a part-time Professor with the University of Brussels (VUB), Brussels, Belgium. He authored or coauthored six books and more than 350 scientific papers. His main interests are in analog, RF, and mm-wave IC design for wireless communication and sensing applications.

Dr. Wambacq has been on the Program Committee of solid-state circuit design conferences, such as European Solid-State Circuits Conference (ESSCIRC) and International Solid-State Circuits Conference (ISSCC). He was the Vice-Program Chair of ISSCC 2022 and the Program Chair for ISSCC 2023.



Jan Craninckx (Fellow, IEEE) received the M.S. and Ph.D. degrees *(summa cum laude)* in microelectronics from the ESAT-MICAS Laboratories, KU Leuven, Leuven, Belgium, in 1992 and 1997, respectively. His Ph.D. work was on the design of low-phase noise CMOS integrated voltage-controlled oscillator (VCO) and phase-locked loop (PLL) for frequency synthesis.

From 1997 to 2002, he worked with Alcatel Microelectronics (later part of STMicroelectronics), Brussels, Belgium, as a Senior RF Engineer on the integration of RF transceivers for global system for mobile communications (GSM), digital enhanced cordless telecommunications (DECT), Bluetooth, and wireless LAN (WLAN). In 2002, he joined imec, Leuven, Belgium, as a Principal Scientist working on RF, analog, and mixed-signal circuit designs. He is currently an imec fellow. He has authored or coauthored more than 200 papers, book chapters, and patents. His research focuses on the design of RF transceiver front ends in nanoscale CMOS, covering all aspects of RF, mm-wave, analog, and data converter designs.

Dr. Craninckx is/was a Regular Member of the Technical Program Committee for several IEEE Solid-State Circuits Society (SSCS) conferences. He was an elected SSCS AdCom member from 2017 to 2019. He was the Chair of the SSCS Benelux Chapter from 2006 to 2011. He also serves as the RF Subcommittee Chair of the IEEE International Solid-State Circuits Conference. He was an Associate Editor and the Editor-in-Chief of the IEEE JOURNAL OF SOLID-STATE CIRCUITS from 2009 to 2016 and 2016 to 2019, respectively. He was an SSCS Distinguished Lecturer from 2012 to 2013.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 28,2022 at 05:08:53 UTC from IEEE Xplore. Restrictions apply.

Electrical Engineering
https://doi.org/10.1007/s00202-022-01571-y

ORIGINAL PAPER

Design and analysis of Quadratic Bernstein Functional Blending Neural Network for shunt compensation and Phase Locked Loop

Ankita Arora[1] · **Alka Singh**[1]

Received: 19 October 2021 / Accepted: 19 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract
This paper proposes a Quadratic Bernstein Functional Blending Neural Network designed for shunt compensation as well as Phase Locked Loop controller in synchronizing the power electronic converters to the grid-connected systems. It addresses all the current-based power quality problems such as power factor correction, harmonics alleviation, reactive power compensation, and all voltage-based power quality problems under adverse grid conditions such as voltage sag, swell, voltage harmonics, dc offset, frequency, and grid phase angle fluctuations simultaneously. It is designed on the Quadratic Functional Expansion of the input signal using Bernstein Polynomial expansion. The novel feature of this polynomial is that its expansion exhibits maximum differentiable and convex hull property ensuring the smoothness of the estimated load and voltage curve. Further, the technique is also made adaptive and is trained online by regularly updating weight coefficients using Least Mean Square Algorithm. Its simplified structure provides advantage over higher order trained neural networks which show sluggish dynamic response and inherent delay in tracking of estimated quantity. The proposed controller is verified successfully on hardware developed in the laboratory and its performance is validated under various grid and load disturbances.

Keywords Power quality · Shunt compensation · Phase locked loop · Synchronization · Neural network

1 Introduction

High harmonic distortion is caused by increased use of non-linear loads and power electronic devices connected at the point of common coupling (PCC) in power distribution systems which has led to several current and voltage power quality-based problems [1]. These high harmonic distortions and disturbances at PCC voltage cause losses and oscillations leading to the equipment failure. Under non-ideal and weak grid, the presence of grid disturbances such as voltage swell, sag, dc offset, harmonic distortions, frequency, and voltage phase angle variation further add to challenging problems [2]. Moreover, constantly changing load variations with high percentage of non-linear loads have prompted power engineers to devise new solutions [3, 4]. Effective solutions to minimize these power quality problems necessitate good shunt compensation as well as new synchronization techniques for effective control of grid-connected inverters.

✉ Alka Singh
alkasingh.eed@gmail.com

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

Synchronization of the power electronic converters with the grid is mandatory as per new grid codes [5]. This has resulted in accurate measurement of voltage magnitude, phase angle, and frequency. The existence of weak and non-ideal grid in developing countries having fluctuations has made this task even more challenging. Therefore, Phase Locked Loops (PLL) are widely employed for synchronization and tracking of parameters such as frequency and phase angle. Various PLLs are described in the literature [6–8] but the challenge is how to achieve fast and effective response under non-ideal grid variations. The conventional PLLs fail to respond instantaneously under large frequency variation, phase angle jump/dip and the response deteriorates further under polluted grid. Most of the PLLs such as Synchronous reference frame theory (SRFT) [6, 7], Second order generalized Integrator (SOGI) [8, 9] are popular techniques used for synchronization and they work quite well under normal grid conditions but their performance degrades under abnormal grid conditions. SRFT PLL shows high transients and oscillations, large steady state error under phase and frequency variations [6, 7]. SOGI PLL fails to estimate the signal when the grid voltage has DC offset, inter-harmonics and use of integrator further cause delay in the estimation process. The

Published online: 21 May 2022

selection of appropriate gain 'k' is one of the drawbacks of SOGI systems [8–10]. Several other novel PLLs such as Cascaded Delayed Cancellation (CDSC) [11], Fourier Transform [12], Notch filter [13] are frequency-dependent PLLs which are effective but encounter problems in bandwidth selection. Some other PLLs mentioned in recent literature include Adaptive Spline [14] and the Predictive Model Control [15]. Adaptive Spline technique is mathematically intensive and based on Spline basis matrix of higher order. In predictive control PLL, the output response is poor and sluggish as it is based on the previous responses. This hinders fast and effective synchronization, especially under large disturbances.

Shunt compensation is employed for solving current-based power quality problems. Literature review suggests that the Shunt Active power Filter (SAPF) working in current control shunt mode is a very effective technique to inject compensating currents. It performs multiple functions viz. harmonic elimination, reactive power compensation, and also improves power factor of supply current to nearly unity when poor power factor loads are connected at the PCC [16–21]. Nishant et al. [16] have discussed Multiple Fifth order generalised Integrator (MFOGI) based Shunt Active Power Filter (SAPF) which has good filtering capability but due to its higher order, it shows sluggish response. Bagheri et al. and Chittora et al. have discussed Kalman Filtering technique for control of SAPF, however, the selection of Kalman gain is cumbersome and a challenging task [17, 18]. Moreover, extended and recursive versions of Kalman filter have been referred to in the literature but the estimation of various state variables using several equations increases the complexity further [19–21]. Adaptive shunt compensation techniques in literature mention Adaptive observer-based control method [22], Widrow-Hoff based Adaptive Filter [23] but these methods involve large calculations, state space realization and increased overall complexity with marginal benefits only. In recent years, the application of fuzzy logic, Neural Networks, and Adaptive Neuro-fuzzy Inference systems algorithms is also hugely popular [24–27]. Several Neural Network based controllers have been designed with different number of hidden layers, neurons, and activation function, etc. Back-propagation Neural Network (BPNN) controller is a conventional controller designed using the weights updated by error propagated backward [25–27]. There are other recent higher order controllers based on Legendre polynomials [28], Chebyshev polynomials [29], and Adaptive Volterra-based filters [30]. These work well under normal grid conditions but have not been tested under adverse grid conditions thoroughly. The advantage of using higher order polynomials in shunt compensation is immense. They possess capabilities comparable to NNs in terms of fast processing and estimation; moreover, there is no hidden layer requirement.

Therefore, there is a need to design controller which works as an effective PLL for synchronization as well as good shunt compensator. In this paper, Quadratic Bernstein Polynomial expansion has been utilized for designing a shunt compensator and a new synchronization scheme. The expansion exhibits the maximum differentiable and convex hull property of the polynomial ensuring the smoothness of the estimated interpolated curve. It belongs to the category of Functional Link Neural Bernstein Blending Networks [31–34] which belong to the family of Spline curves which are powerful models for non-linear systems. They exhibit spline interpolating properties having values between 0 and 1 and for any degree, all the Bernstein coefficients sum up to one. Therefore, the challenge to select the differentiable activation function is altogether eliminated. The application of Quadratic Bernstein Functional Blending Neural Network (QB-FBNN) has been discussed for noise filtering in communication systems such as satellite, microphone, hearing aids, and echo cancellation but has not been exploited in power systems for shunt compensation and synchronization to date. The important contributions of this controller are:

1. QB-FBNN belongs to the Family of Spline interpolating non-linear curves. This paper highlights that their effective design can be realized both for synchronization and load compensation.
2. The controller developed works as a second order filter which is found to be effective for both linear and non-linear loads.
3. The computational burden of the designed controller is not very high; thus the control is fast, effective, and can be implemented using low-cost digital signal processors.
4. Its fast response avoids unwanted high dip in dc link voltage and fast settling of dc link voltage to its reference value, thus playing an important role in stabilizing the DC link voltage under large disturbances.
5. The controller is tested successfully for both load compensation and as a PLL for synchronization which is an added advantage.
6. The self-adaptive nature of the controller avoids the need to tune it under several disturbances. Good performance is guaranteed under adverse grid parameters such as voltage sag, voltage swell, harmonics and, especially phase angle and frequency disturbance occurring simultaneously.
7. Challenges in the selection of differentiable activation function are easily overcome due to unique properties of these functions.

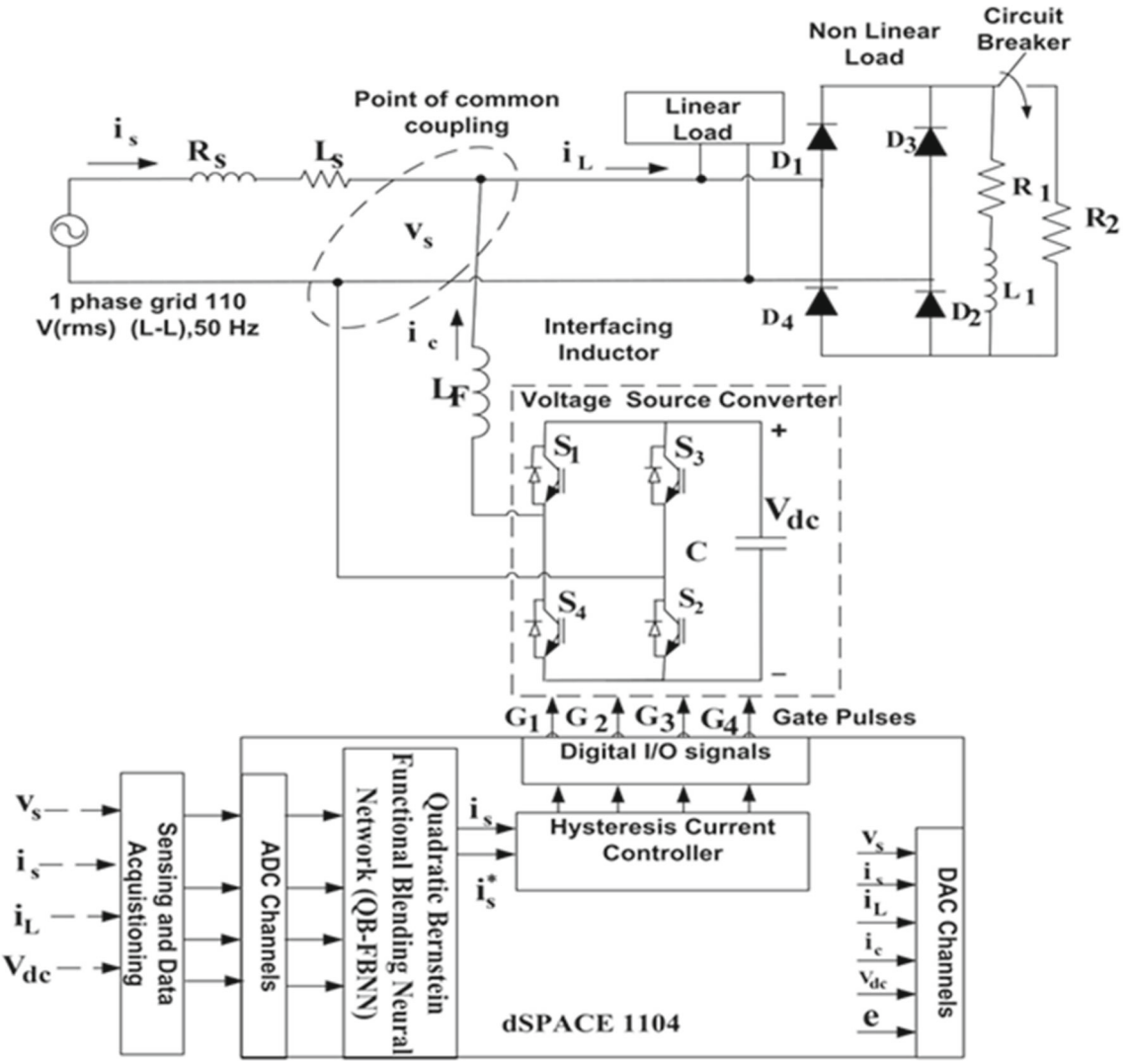


Fig. 1 Schematic block diagram and controller for Quadratic Bernstein Functional Blending Neural Network

2 Proposed configuration

Figure 1 shows the single-phase grid-connected system supplying linear and non-linear loads. The voltage source converter (VSC) in combination with the capacitor is referred to as Shunt Active Power Filter (SAPF). The SAPF is designed and controlled to inject compensating current to cancel the harmonics at the PCC injected by the power electronic converters and non-linear load connected to it. The interfacing inductor (L_F) is designed and installed to minimize the ripples in the compensating current as well as noise generated due to switching of SAPF. The VSC is controlled using QB-FBNN algorithm. Synchronization is implemented with the help of proposed algorithm acting as a PLL to generate the sinusoidal reference signals under adverse grid conditions. The paper discusses the performance under several cases involving both normal and adverse grid conditions. These include voltage sag, voltage swell, harmonic distortion, frequency and phase angle variation, dc offset, and sudden load changes. The next section discusses both the aspects of compensation and synchronization in detail.

3 QB-FBNN controller as shunt compensator and PLL

3.1 Mathematical formulation of QB-FBNN controller as shunt compensator for estimation of active power load component

QB-FBNN is a Functional Link Neural Network (FLNN) controller designed to work using the method of interpolation. Quadratic Bernstein Functional Curve is formulated

from polynomial expansion terms passing through the first and the last control point i.e. i_{L0} and i_{L2} respectively and contained inside the control polygon called the convex hull of all the three control points i.e. i_{L0}, i_{L1} and i_{L2}. This property ensures that the polynomial smoothly follows the path of the control points. Quadratic Bernstein curve is the path traced by the function $i_{LF}(t)$ consisting of 'm' terms given by the expression defined below:

$$i_{LF}(t) = \sum_{k=0}^{m} i_{Lk} B_{m,k}(t) \tag{1}$$

where Bernstein Coefficient $B_{m,k}(t)$ is defined as

$$B_{m,k}(t) = \binom{m}{k}(1-t)^{m-k}t^k \tag{2}$$

such that $\sum_{k=0}^{m} B_{m,k}(t) = 1$ unlike other higher order polynomials

$$i_{LF}(t) = \sum_{k=0}^{m} i_{Lk}\binom{m}{k}(1-t)^{m-k}t^k \tag{3}$$

where $\binom{m}{k} = \frac{m!}{(m-k)!k!}$ which can be expressed recursively as

$$i_{LF}(t) = (1-t)\left(\sum_{k=0}^{m-1} i_{Lk} B_{m-1,k}(t)\right) + t\left(\sum_{k=1}^{m} i_{Lk} B_{k-1,m-1}(t)\right) \tag{4}$$

The above expression is a linear interpolation between two-degree $m-1$ Bernstein curves. For quadratic Bernstein Curve we have $m = 2$,

$$i_{LF}(t) = \sum_{k=0}^{2} i_{Lk} B_{2,k}(t) \quad k = 0,\ 1 \ldots m \tag{5}$$

$$i_{LF}(t) = i_{L0}B_{2,0}(t) + i_{L1}B_{2,1}(t) + i_{L2}B_{2,2}(t) \tag{6}$$

$$i_{LF}(t) = (1-t)^2 i_{L0} + 2t(1-t)i_{L1} + t^2 i_{L2} \tag{7}$$

$$i_{LF}(t) = \left(1 + t^2 - 2t\right)i_{L0} + 2\left(t - t^2\right)i_{L1} + t^2 i_{L2} \tag{8}$$

$$i_{LF}(t) = i_{L0} + t(-2i_{L0} + 2i_{L1}) + t^2(i_{L0} - 2i_{L1} + i_{L2}) \tag{9}$$

$$i_{LF}(t) = \begin{bmatrix} 1 t t^2 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ -2 & 2 & 0 \\ 1 & -2 & 1 \end{bmatrix} \begin{bmatrix} i_{L0} \\ i_{L1} \\ i_{L2} \end{bmatrix} \tag{10}$$

Generalizing the above matrix, we get

$$i_{LF}(\mathrm{t}) = \Big[\left(t^2 - 2t + 1\right)i_{L(k-2)}, \left(-2t^2 + 2t + 1\right)i_{L(k-1)},\ t^2 i_{L(k)}\Big] \tag{11}$$

where

$$i_{LF} = [i_{LF0}(\mathrm{t}),\ i_{LF1}(\mathrm{t}),\ i_{LF2}(\mathrm{t})] \tag{12}$$

Therefore, the curve consists of three computed expansion components i_{LF0}(t), i_{LF1}(t),i_{LF2}(t). Also, it is highly desired that the entire process is self-learning in nature and the computation of current evaluation is automatic. This feature is incorporated by computing the weight coefficient 'w' iteratively in each expansion component and adding all the components to calculate the estimated load current expressed as $i_{Lest}(t)$. The controller is designed to extract only the active power component of the load current. Figure 2a shows the detailed control diagram designed for shunt compensation.

$$i_{Lest} = [wi_{LF0}(t) + wi_{LF1}(t) + wi_{LF2}(t)] \tag{13}$$

The estimated load current computed by QB-FBNN based controller for shunt compensation is given by (14)

$$i_{Lest}(t) = w\Big[i_L^2(k-2) - 2i_L(k-2) + i_L(k-2) - 2i_L^2(k-1) + 2i_L(k-1) + i_L(k-1) + i_L^2(k)\Big] \tag{14}$$

The controller is made adaptive by using Least Mean Square (LMS) technique. The error (e) between the actual load current and estimated load current is computed and minimized by updating the weights coefficients 'w' iteratively in real time by the learning algorithm.

$$e = i_L(t) - i_{Lest}(t)\sin\theta \tag{15}$$

where $\sin\theta$ is the synchronization template obtained from the QB-FBNN PLL. Figure 2b shows the proposed scheme realized as a PLL and the inputs include the voltage components and the output is the computed template $\sin\theta$.

The product of i_{Lest}(t) with $\sin\theta$ ensures the fundamental current component is considered for control purpose. This term represents the grid current that needs to be supplied for achieving unity power factor operation. The weight updating is discussed in (16–17).

$$w(k+1) = w(k) - \frac{\mu}{2}\frac{\partial E\left(e^2\right)}{\partial i_{Lest}} \tag{16}$$

$$w(k+1) = w(k) + \mu e \sin\theta \tag{17}$$

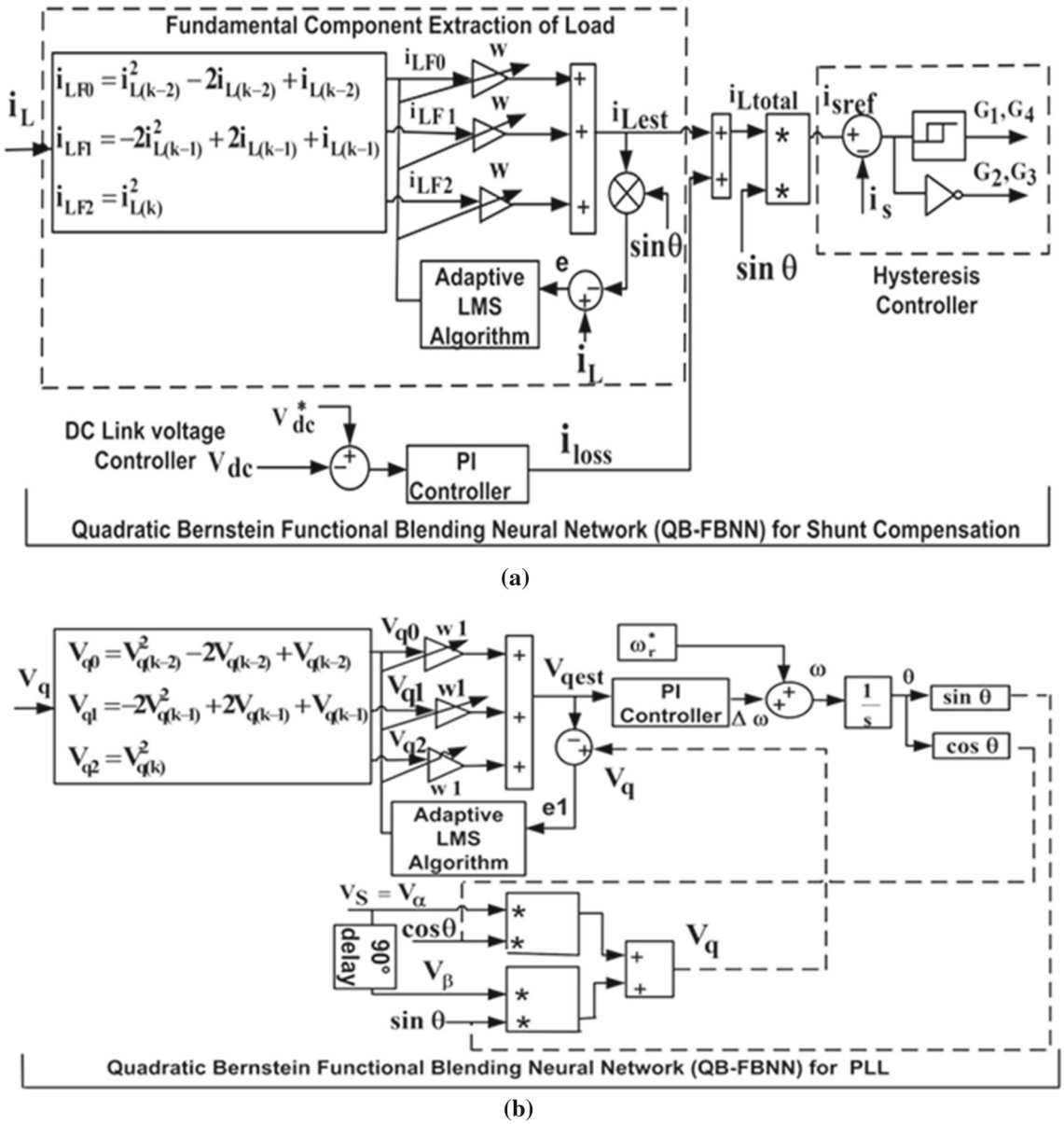


Fig. 2 Quadratic Bernstein Functional Blending Neural Network controller in grid connected system for **a** shunt compensation and **b** synchronization as Phase Locked Loop

where μ indicates the learning parameter, e indicates the error and E indicates the expected mean error.

3.1.1 Computation of reference grid current and pulse generation for VSC

The dc link voltage at the output of the VSC converter is to be regulated for reactive power compensation. The dc link voltage (V_{dc}) at the DC Link of VSC is sensed and compared with the reference dc link voltage (V^*_{dc}). The error ($V_{de}(k)$) between the two is minimized by passing through PI controller which regulates the dc link voltage to its reference value. This output of PI controller is considered as the switching loss component and denoted by i_{Loss}

$$V_{de}(k) = V^*_{dc} - V_{dc} \tag{18}$$

$$i_{Loss}(k) = i_{Loss}(k-1) + K_p\{V_{de}(k) - V_{de}(k-1)\} + K_i V_{de}(k) \tag{19}$$

The controller is designed so that the total active component of current ((i_{Ltotal}) comprising estimated load current (i_{Lest}) and the switching loss component (i_{Loss}) is drawn from the grid.

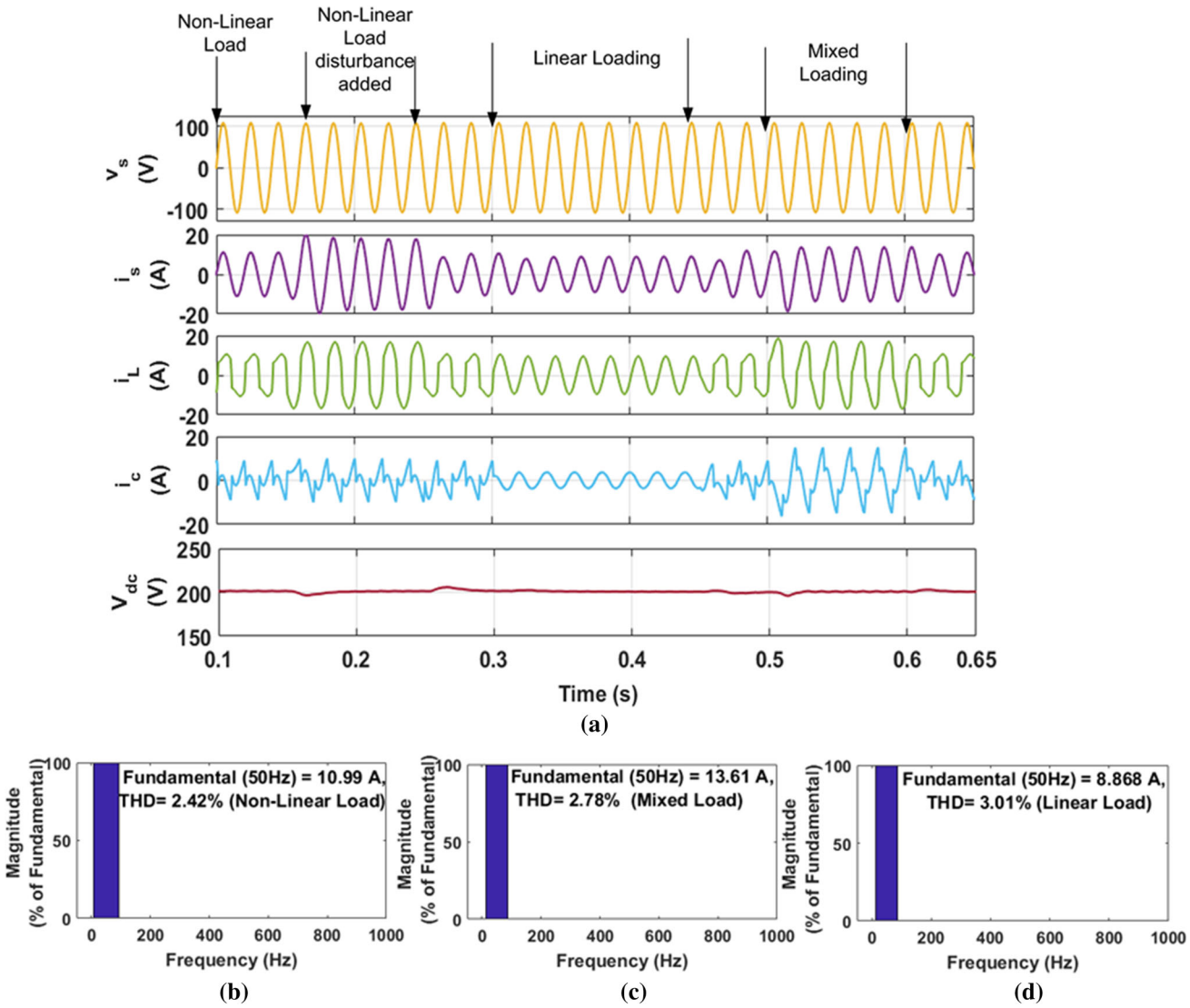


Fig. 3 Performance analysis of QB-FBNN with load variations. **a** Shunt compensation results, **b** THD of grid current under non-linear load conditions, **c** THD of grid current under mixed loading conditions and **d** THD of grid current under linear loading conditions

$$i_{\text{Ltotal}} = i_{\text{Lest}} + i_{\text{Loss}} \tag{20}$$

The reference component of grid current (i_{sref}) is obtained by multiplying i_{Ltotal} with the unit-in phase template

$$i_{\text{sref}} = i_{\text{Ltotal}} \sin\theta \tag{21}$$

Here, $\sin\theta$ is the unit in phase template generated by the proposed controlled which works as a PLL. The reference grid current is compared with the actual source current using hysteresis current controller to produce appropriate switching pulses for the VSC for shunt compensation. The hysteresis band of the controller is selected to be 0.2.

3.2 QB-FBNN controller working as a PLL

The grid voltage (v_s) signal is fed to the synchronization PLL block which takes the input in $\alpha\beta$ coordinates, v_α voltage is the same as supply voltage (v_s) and v_β is obtained by incorporating a phase shift of $\frac{\pi}{2}$ radians in supply voltage. These $v_{\alpha\beta}$ components are converted to dq reference frame voltages through Park transformation shown below.

$$\begin{bmatrix} V_d \\ V_q \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} v_\alpha \\ v_\beta \end{bmatrix} \tag{22}$$

where θ is the grid phase angle computed and updated iteratively in closed loop operation. The adaptive QB-FBNN PLL uses V_q as its input to the controller and regulates its value to

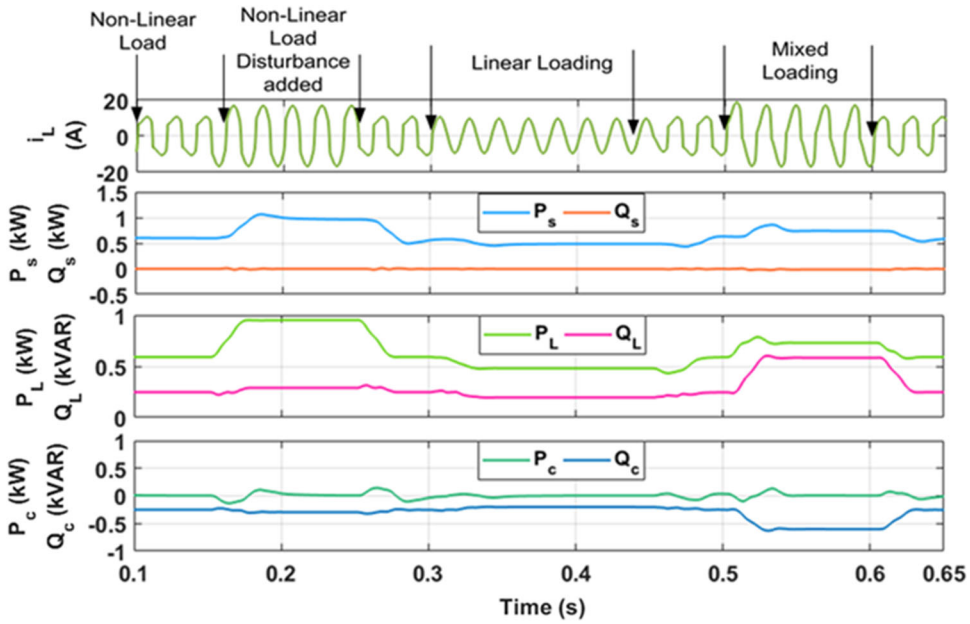


Fig. 4 Performance analysis of steady state power waveforms of QB-FBNN with load variations

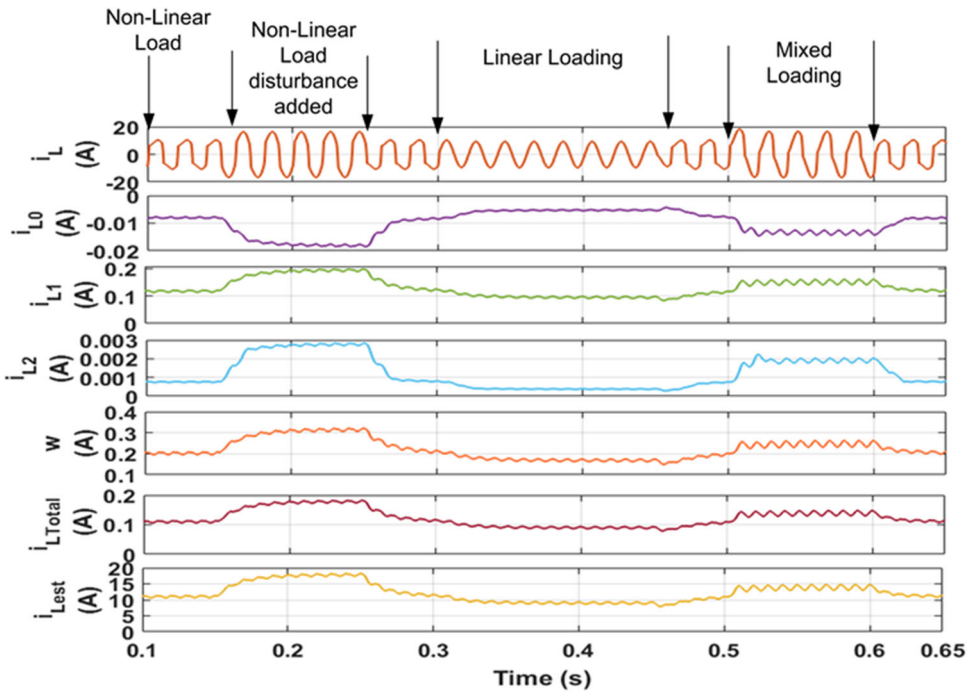


Fig. 5 Weight convergence using the proposed controller showing several terms

zero irrespective of changes in frequency, phase angle, voltage dip, voltage rise, dc offset, and harmonic distortion in weak grids. The designed controller computes V_{qest} with the help of three expansion components.

$$V_{qest}(t) = w_1\Big[V_q^2(k-2) - 2V_q(k-2) + V_q(k-2), \\ -2V_q^2(k-1) + 2V_q(k-1) + V_q(k-1),\ V_q(k)\Big] \quad (23)$$

It implements the adaptive closed loop control process which minimizes the error between the reference quadrature voltage V_q and the estimated quadrature voltage V_{qest} using Least Means Square (LMS) algorithm by updating its weight coefficient $'w_1'$. The control of reactive voltage component (V_{qest}) to zero is useful in regulating the grid parameters such as voltage, frequency, dc offset and phase angle oscillations to its actual value.

$$E(V_{sm}, \theta) = \frac{1}{2}e_1^2 = \frac{1}{2}\big(V_q - V_{qest}\big)^2 \quad (24)$$

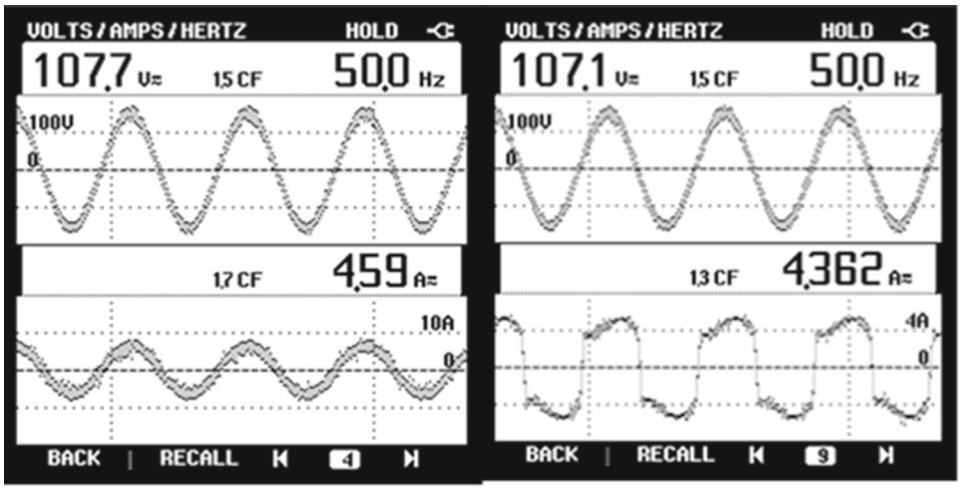


Fig. 6 Steady state waveform **a** source current, **b** load current, **c** source current THD and **d** load current THD

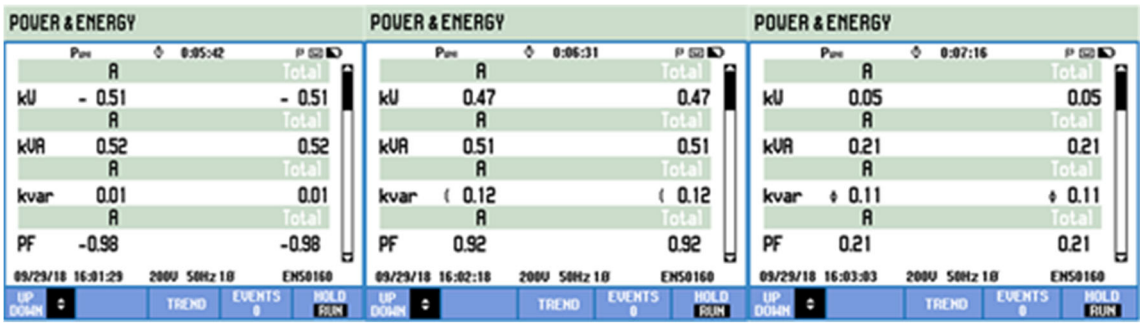


Fig. 7 Steady state power waveform **a** source power, **b** load power and **c** compensator power

This error computed in (24) is minimized by updating the weight in real time by LMS Algorithm till it gets converged.

$$w_1(k+1) = w_1(k) - \frac{\mu}{2}\frac{\partial E_1\left(e_1^2\right)}{\partial V_{qest}} \tag{25}$$

$$w_1(k+1) = w_1(k) + \mu e_1 \sin\theta \tag{26}$$

Figure 2b shows the detailed diagram where the V_{qest} obtained is applied as input to the PI controller whose output provides the error in angular frequency $(\Delta\omega)$. This angular frequency error $(\Delta\omega)$ is added to the natural frequency $\left(\omega_r^*\right)$ to get the estimated angular frequency.

$$\omega_{\text{est}} = \omega_r^* + \Delta\omega \tag{27}$$

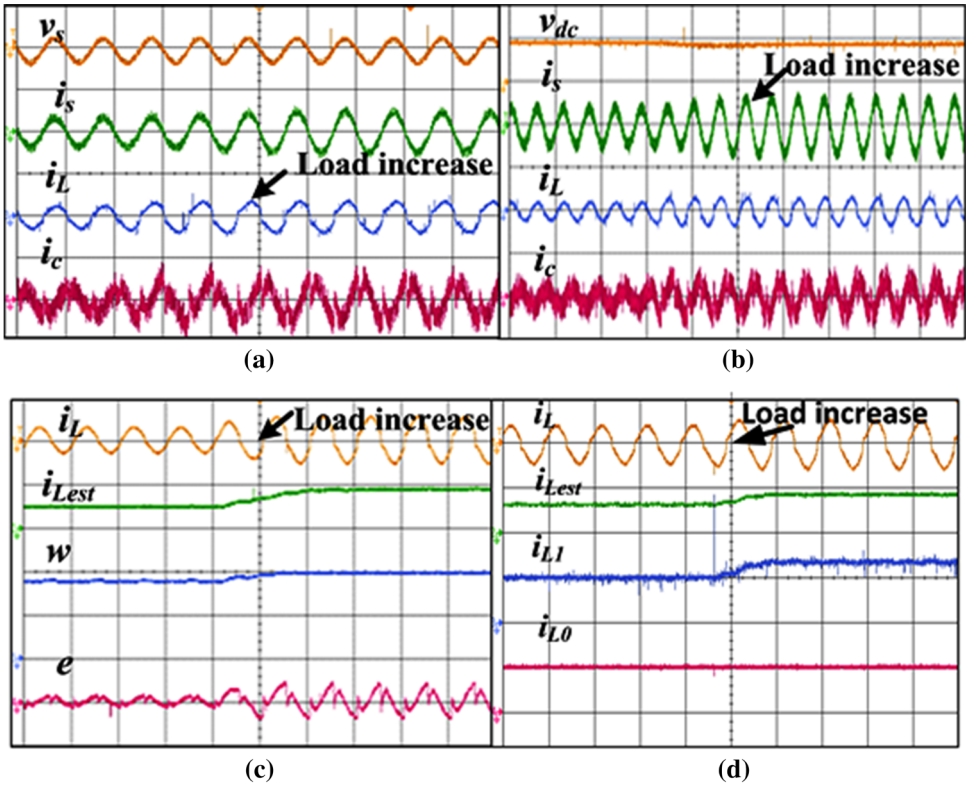


Fig. 8 Dynamic observed under increased load conditions on oscilloscope **a** v_s, i_s, i_L, i_c, **b** v_{dc}, i_s, i_L, i_c, **c** i_L, i_{Lest}, w, e, **d** i_L, i_{Lest}, i_{L0}, i_{L1}

The estimated frequency (ω_{est}) is integrated to get the grid voltage angle (θ_{est}).

$$\int \omega_{est} dt = \theta_{est} \tag{28}$$

The estimated voltage angle (θ_{est}) is used to get the unit in-phase template and the quadrature template for generation of reference signal.

4 Performance analysis as shunt compensator and PLL

4.1 Performance analysis as shunt compensator

The QB-FBNN controller is developed using Simulink and the performance of shunt compensator is observed under different operating conditions such as mixed loading, non-linear loading, and linear loading conditions. The performance is analyzed with various response parameters such as grid voltage (v_s), grid current (i_s), load current (i_L), compensator current (i_c), DC Link voltage (V_{dc}), and source Power (P_s), load Power (P_L), compensator power (P_c) are shown in Figs. 3 and 4 respectively.

Under normal non-linear load conditions (t = 0–0.15 s), the load consumes 600 W, which is met by the source. The dc link voltage is regulated to 200 V. The load reactive power of 250 VAR is met completely by the compensator. Non-linear load is increased from 0.15 to 0.25 s. The source current increases to 16.5 A to supply 950 W to meet the additional load demand. The source current is observed to be sinusoidal with the grid voltage indicating unity power factor operation. The THD of the source current under non-linear load conditions is observed to be 2.42% when the non-linear load has THD of 30%.

Under linear loading conditions from 0.3 to 0.45 s, the source current, load current, and compensator current all are sinusoidal. The source current of 9.5 A is drawn to meet the load demand of 480 W and the reactive load of 200 VAR is met by the compensator.

Under mixed loading conditions comprising both linear and non-linear load, the source current is sinusoidal with a peak value of 14 A to meet the load demand of 732 W. The reactive power demand of 586 VAR is met by the compensator. The THD under mixed loading conditions is observed to be 2.78%.

Figure 5 illustrates the intermediate results of the control algorithm consisting of load current, estimated value, weight (i_L, i_{Lest}, 'w') and individual expansion components i_{L0}, i_{L1},i_{L2}, and i_{Ltotal}. It is found that the first two expansion components are sufficient in estimating the load current. The proposed QB-FBNN technique can be further simplified by considering the dominant terms contributing effectively.

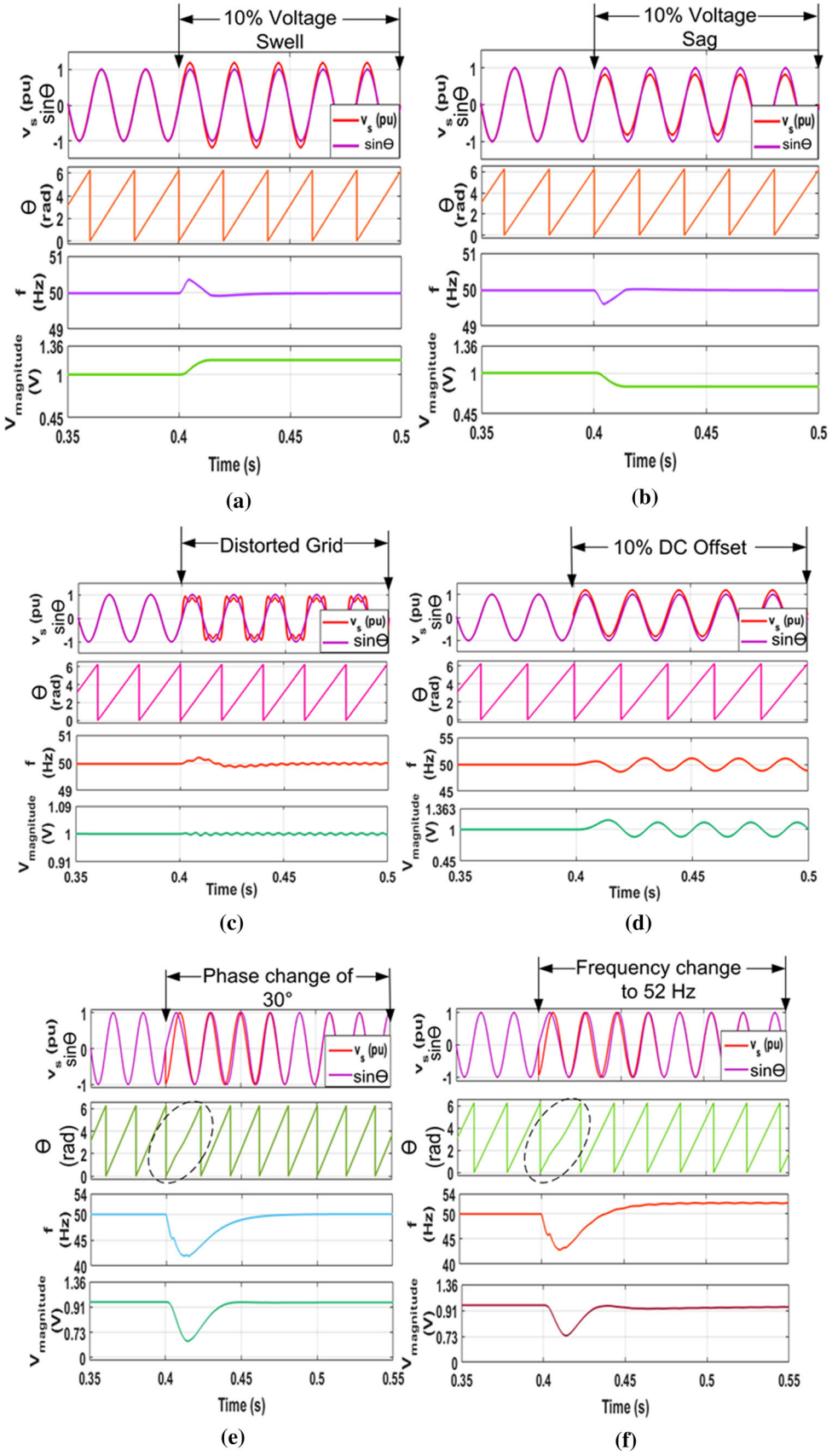


Fig. 9 Performance of QB-FBNN algorithm as PLL under various grid conditions **a** voltage swell of 10%, **b** voltage sag of 10%, **c** distorted grid, **d** 10% DC offset, **e** phase change of 30° and **f** frequency change to 52 Hz

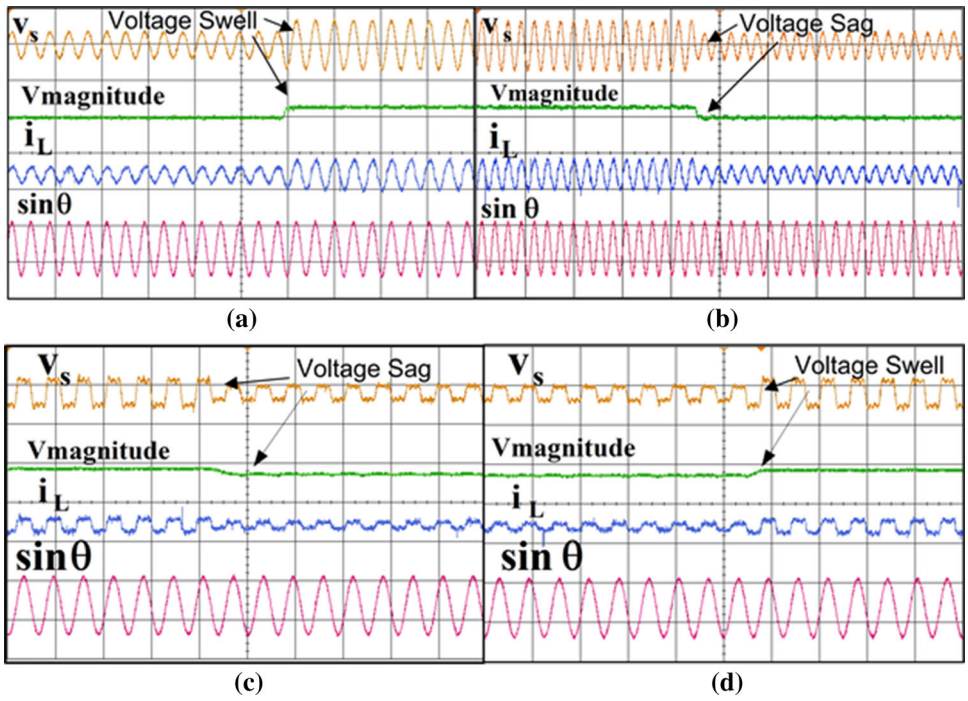


Fig. 10 Experimental dynamic results for QB-FBNN PLL under **a** voltage swell, **b** voltage sag, **c** distorted grid with voltage swell and **d** distorted grid with voltage sag

Figure 6 illustrates the steady state waveform and THD of source current and load current using hardware results. These results illustrate that even under highly non-linear load having THD of 30% observed in Fig. 6d, the QB-FBNN controller removes the harmonics injected into the source current and improves the THD to 2.9% shown in Fig. 6c. The source current is observed to be sinusoidal and in phase with the source voltage.

Figure 7 illustrates the steady state power balance recorded on the hardware system developed in the laboratory. The power results shown in Fig. 7 shows that out of 510 W power demanded from the grid, 470 W is consumed in the load, and remaining is utilized in meeting the switching losses. The reactive power load demand of 120 VAR is met almost completely by the compensator (110 VAR) and source supplies (10 VAR). The source power factor has improved to 0.98. The THD of supply current is around 2.9% and is sinusoidal as shown in Fig. 7c.

These results justify the control scheme which is designed and implemented to achieve load compensation effectively. Almost the entire reactive power demand is met from the compensator only, thereby achieving a power factor close to unity at the supply side.

4.1.1 Dynamic results

Figure 8 shows the dynamic results. Figure 8a shows the waveforms of v_s, i_s, i_L and i_C. The source current is observed to be sinusoidal and as the load is increased, the source current also increases. Figure 8b shows the variation of dc link voltage with load. As the load is increased, a little dip in dc link voltage is observed but the PI controller regulates it. Figure 8c shows the variation of estimated load current i_{Lest}, error (e) in current estimation and individual load current i_L and fundamental weight component (w). Figure 8d shows the variation of estimated load current and individual load expansion components with load decrease. It has been observed from both the simulation and dynamic results that only first two expansion terms i_{L0}, i_{L1} are sufficient in estimating the load current. This also justifies that a simplified structure of the network is sufficient to achieve enhanced performance.

4.2 Performance analysis as a PLL under various grid perturbations

The performance evaluation of QB-FBNN as PLL is discussed under weak grid parameter variations such as voltage sag, swell, dc offset, harmonics, phase and frequency change. The analysis of waveforms of various grid parameters v_s, i_s, i_L, θ, V_{DC}, $\sin\theta$ and frequency f is illustrated in Fig. 9. The performance of the proposed algorithm is observed under wide variations in grid voltages.

- *Case I* Voltage swell of 10% in grid voltage
  The voltage swell of 10% is simulated in grid voltage shown in Fig. 9a. It has been observed that the voltage amplitude estimated increases showing voltage swell. The frequency shows a minute rise of 0.2 Hz and settles down

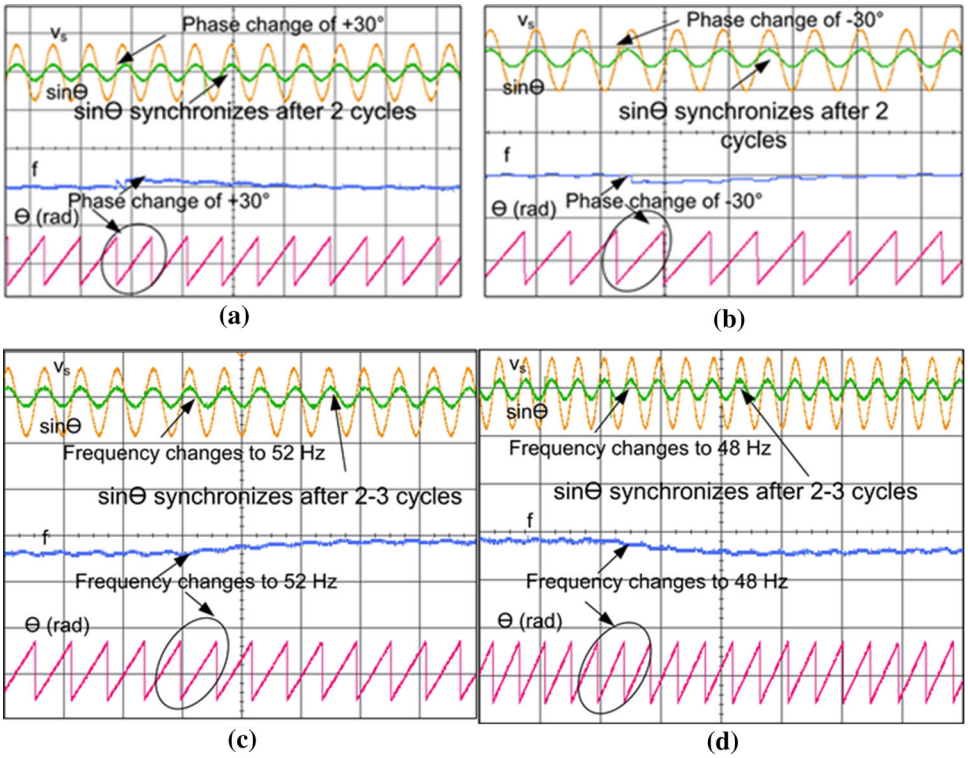


Fig. 11 Experimental dynamic results for QB-FBNN PLL under **a** phase change of + 30°, **b** phase change of − 30°, **c** frequency change of + 2 Hz i.e. 52 Hz and **d** frequency change of − 2 Hz i.e. 48 Hz

to 50 Hz again within 1 cycle. The synchronizing signal sinΘ is sinusoidal and in phase with the supply voltage.

- *Case II* Voltage sag of 10% in grid voltage
  The voltage sag of 10% is simulated in grid voltage shown in Fig. 9b. It has been observed that the voltage amplitude estimated decreases showing voltage sag properly. The frequency shows a little dip of 0.2 Hz and settles down to 50 Hz again within 1 cycle. The synchronizing signal sinΘ is sinusoidal and in phase with the supply voltage.
- *Case III* Under distorted grid voltage
  The grid voltage with THD of 20% having high harmonic content is shown in Fig. 9c. The estimated synchronizing signal sinΘ is sinusoidal and in phase with the voltage. The voltage magnitude and frequency estimated show minimal oscillations under distorted grid.
- *Case IV* DC offset of 10% in grid voltage
  The voltage shows 10% dc offset in grid voltage shown in Fig. 9d. The estimated synchronizing signal sinΘ is sinusoidal and in phase with the voltage. The voltage magnitude and frequency tracked show perturbations during the transient.
- *Case V* Phase change of 30° in grid voltage
  The phase change of 30° is introduced in grid voltage in Fig. 9e. The estimated frequency shows a dip of 5 Hz and voltage estimated shows a dip of 0.18 pu. The synchronizing template sinΘ deviates initially and then correctly estimates and comes in phase with the voltage within 2–3 cycles. This shows the effectiveness of the designed PLL which works effectively.
- *Case VI* Frequency change of 2 Hz in grid voltage

The frequency change of 2 Hz is introduced in grid voltage in Fig. 9f. A dip in frequency and voltage magnitude is observed. The synchronizing template tracks the voltage correctly within 2–3 cycles. However, the new frequency is tracked effectively showing that the proposed QB-FBNN controller works effectively as a PLL also.

4.2.1 Dynamic results

This Section shows various results under sudden perturbations in utility grid.

Performance analysis under normal voltage sag and voltage swell, distorted voltage sag and voltage swell Figure 10 shows the performance of QB-FBNN as PLL under voltage swell and sag in normal and distorted grid conditions. Figure 10a, b shows that the pure sinusoidal unit template is obtained in both the cases. The voltage amplitude $\left(V_{\text{magnitude}}\right)$ and i_{L} increase and decreases proportionately with the normal voltage swell (~ 10%) and sag (~ 10%) respectively. Figure 10c, d show the results when the grid voltage is polluted with 3rd and 5th order harmonics and in this case, also, pure sinusoidal sine template waveform is obtained from the designed PLL. Thus, the performance of QB-FBNN PLL

Table 1 Comparison of QB-FBNN based filter for shunt compensation with other conventional techniques

S. no	Parameter	QB-FBNN	BPNN	SRFT
1	Type	Adaptive	Adaptive	Not adaptive
2	Convergence rate	Fast	Very slow	Medium
3	Convergence of weights			
	Steady state oscillations	Almost negligible	Sustained oscillations of medium amplitude	Sustained oscillations of high amplitude
	Transient oscillations	No peak observed	High peak converges within 2 cycles	Low peak converges within 1 cycle
5	Complexity in implementation	Medium	High	Medium
6	PLL requirement	No	No	Yes
7	Settling time (s)	0.01	0.1	0.03
8	Amount of overshoot (V)	Nil	3 V	1 V
9	THD in source current using simulink (%)	2.42	3.34	4.26

under normal conditions and distorted grid is quite satisfactory under voltage sag and swell conditions.

Performance analysis under voltage swell/sag and frequency change and phase change Figure 11 shows the performance analysis of the PLL under simultaneous change in voltage and frequency change and phase change. The waveforms of v_s, f, sin θ, cos θ have been captured on the oscilloscope. Figure 11a, b show two cases viz. phase change of + 30° and − 30°. It has been observed that the frequency deviates a little but settles down soon showing zero steady state error. The sin θ deviates from grid voltage under sudden phase angle change but gets synchronized to it after two cycles. The phase angle plot also shows minor deviation but tracks the correct angle appropriately.

Figure 11c, d show the waveforms of v_s, θ, f, sin θ under frequency change of + 2 Hz i.e. 52 Hz and − 2 Hz i.e. 48 Hz respectively. When the frequency change takes place, it leads to variation in frequency and phase angle but they soon settle down to their reference value. The obtained sine template is synchronized to the grid voltage after 2–3 cycles. The frequency is also correctly tracked in both cases within 2–3 cycles.

5 Comparison analysis as shunt compensator and PLL

5.1 Comparison analysis as shunt compensator

The method surpasses conventional methods of fundamental component extraction. This is clear from weight convergence comparison of QB-FBNN with one conventional and adaptive algorithm. The weight convergence is the fundamental component extracted from load current. The testing

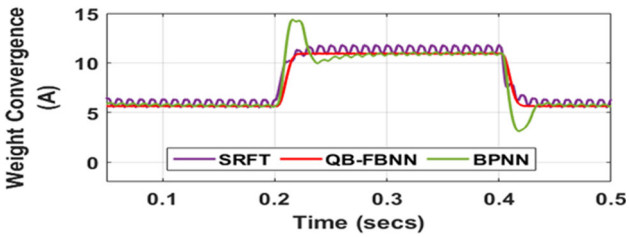


Fig. 12 Weight convergence with different controllers

is performed under changing load conditions/ dynamics to establish the superiority of the method. For this purpose, the fundamental weight of the load current is estimated using the proposed QB-FBNN algorithm and compared with conventional time domain technique SRFT algorithm and the adaptive Neural Network algorithms such as Back-propagation Neural Network (BP-NN). Very fast tracking of weight with no overshoot and negligible oscillations are noticed with QB-FBNN technique.

A detailed comparison has been carried out on the basis of various parameters such as type of algorithm, weight convergence rate, THD in source current, complexity in implementation, settling time, time required to achieve convergence in fundamental weight estimation of load currents during the transient and steady state condition. These results have been compiled in Table 1.

Further, Fig. 12 shows the comparison of weights using three different algorithms. The weight convergence of QB-FBNN is found to be fastest, SRFT medium while BPNN is observed to respond the slowest. QB-FBNN algorithm shows almost no oscillations in tracking load change and achieves convergence in < 1 cycle while BPNN converges in more than two cycles respectively. A major observation of Fig. 12 results show that the QB-FBNN technique showed almost

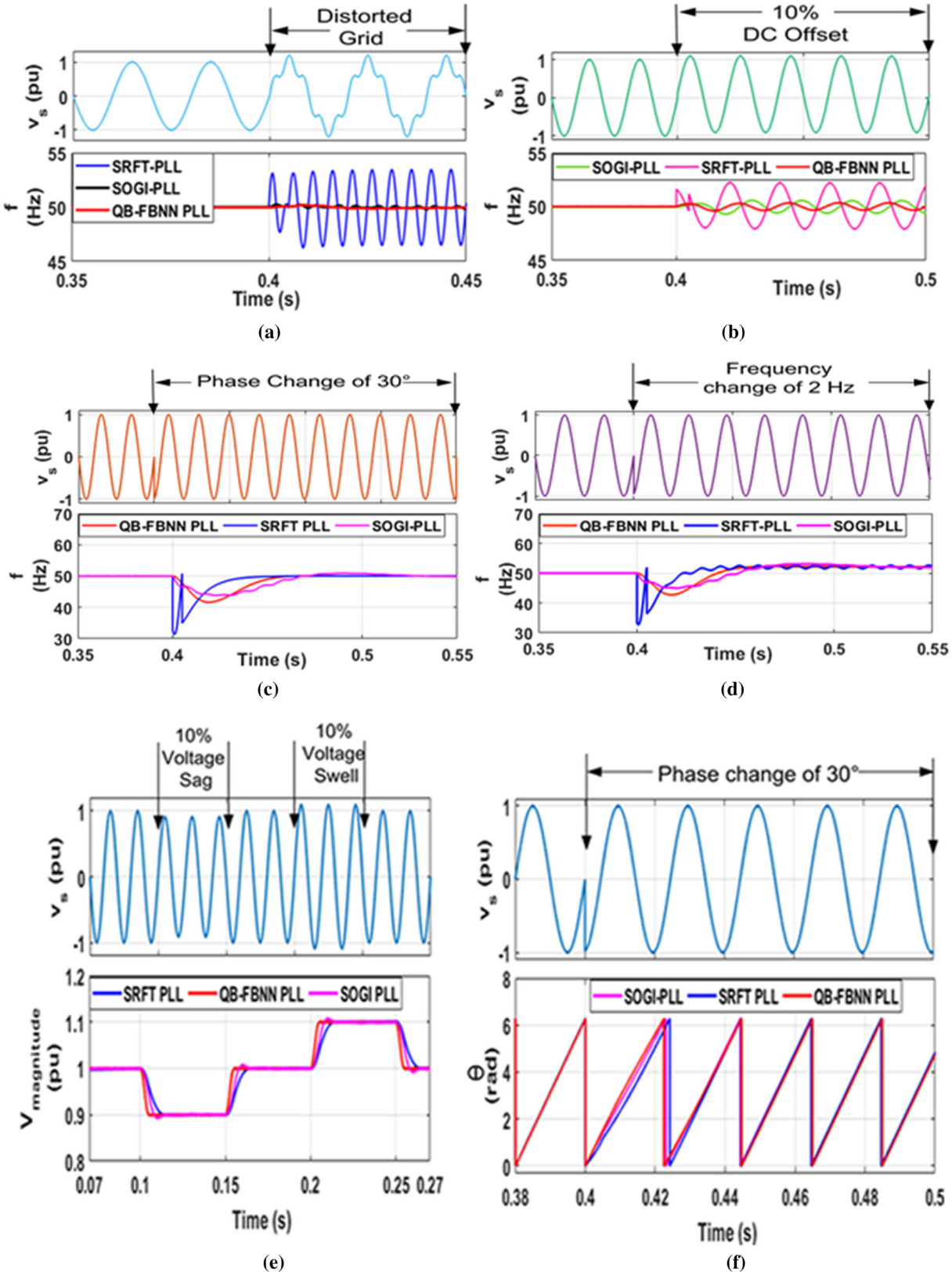


Fig. 13 Comparison analysis of PLLs : **a** 20% distorted grid **b** 10% DC offset **c** frequency changes wrt. phase change of 30 degree **d** frequency change of 2Hz **e** 10% voltage sag and 10% voltage swell **f** phase angle characteristics during Phase change of 30 degree

Table 2 Comparison of different PLL techniques

S. no	Parameter	QB-FBNN	SOGI	SRFT
1	Complexity	Medium	Low	Medium
2	Type	Adaptive	Non-adaptive	Non-adaptive
3	Time taken for amplitude estimation under voltage sag and swell	0.003 s	0.01 s	0.01 s
4	Oscillations in frequency in distorted grid	Negligible amplitude oscillations	Minimum amplitude oscillations	High amplitude sustained oscillations
5	Oscillations in frequency during dc offset	Minimum amplitude oscillations	Minimum amplitude oscillations	High amplitude sustained oscillations
6	Oscillations in frequency during frequency change and phase change	Minimum amplitude oscillations	Minimum amplitude oscillations	High amplitude oscillations
7	Overshoot/undershoot in frequency characteristics during phase change	Medium	Medium	High
8	Time taken to track phase angle change characteristics during phase change	Fast	0.03 s	0.04 s

negligible oscillations, SRFT shows sustained high amplitude oscillations whereas BPNN shows medium amplitude oscillations. Thus, an additional requirement of low pass filter is altogether eliminated for the proposed QB-FBNN controller. It shows best tracking performance under steady state as well as dynamic conditions.

Further, the computational complexity of an algorithm is computed through mathematical blocks such as multiplication, addition, subtraction, and division involved computation. It is found QB-FBNN and SRFT show medium complexity whereas BPNN is the most complex. It is also observed that the performance of SRFT is mainly regulated by the gain constants or PI parameters which need proper tuning. The proposed QB-FBNN technique being self-adaptive in nature automatically adjusts the weights as per the system variations, thus assuring optimum performance under widely varying operating conditions. The THD in source current observed under similar grid and non-linear load conditions are found to be 2.42% with designed QB-FBNN algorithm. Other algorithms such as BPNN, SRFT show THD of 3.34%, 4.26% in source current respectively. Therefore, it can be inferred QB-FBNN performs well in terms of fast convergence under steady state and transient conditions, settling time, low steady state error, and overshoot than other conventional and adaptive algorithms considered for comparison.

5.2 Comparison analysis of different PLLs

Figure 13 shows the comparison results for three PLLs viz. SOGI-PLL, SRFT-PLL and proposed QB-FBNN-based PLL. Figure 13a shows the harmonics in grid voltage having THD of around 20%. The SRFT PLL has high oscillations in frequency in distorted grid, SOGI also has oscillations in frequency and QB-FBNN PLL being adaptive by LMS algorithm has negligible oscillations. Hence SRFT PLL shows poor performance.

Figure 13b shows the grid voltage with 10% offset. Frequency estimation is observed in all the three cases. It has been found that QB-FBNN PLL and SOGI PLL show minimal oscillations and the SRFT shows high amplitude oscillations.

Figure 13c shows the phase change of 30°, high spikes are observed in frequency estimation with SRFT PLL; whereas QB-FBNN and SOGI-PLL show comparatively lower peak

Figure 13d shows the frequency change of 50–52 Hz, it has been observed that SRFT PLL shows continuous sustained oscillations during this duration. It has been found that QB FBNN PLL and SOGI PLL show minimal oscillations.

Figure 13e shows the amplitude tracking ($V_{\text{magnitude}}$) and the time taken by PLLs to settle down to steady state value. It has been observed that the QB-FBNN PLL is fast and responds within 0.003 s which corresponds to less than half a cycle for tracking. But SOGI PLL shows a little peak in starting. SOGI PLL and SRFT PLL take 0.01 s to settle down under both voltage sag and swell.

Figure 13f shows that the QB-FBNN tracks the phase angle instantaneously while SOGI takes 0.03 s and SRFT takes 0.04 s to track phase angle changes. All the results have been recorded in Table 2.

6 Conclusions

The Quadratic Bernstein Functional Blending Neural Network (QB-FBNN) has been designed for both shunt compensation and grid synchronization. It has been tested extensively for achieving both shunt compensation and grid synchronization under various test conditions. Comparison with the conventional algorithms has also been discussed in detail which proves its effectiveness. The results for shunt compensation have shown successfully improved power factor, reactive power compensation, power balance, and harmonic filtering under various grid disturbances. The results obtained meet stipulated IEEE-519 standards for power quality. The QB-FBNN PLL is designed to estimate grid phase angle, frequency change, under sag and swell conditions as well as distorted grid conditions. Its self-adaptive nature makes the controller fast and avoids tuning it under several disturbances making the synchronization control effective. Further, the proposed technique is self-adaptive in nature and works well under wide range of load changes. Its performance as a PLL is a unique feature of the paper and thus the proposed QB-FNN controller will be able to work as effective compensator under distorted and weak grid conditions. The experimental results obtained in both cases match the findings of the simulation results.

Appendix

Simulation parameters

Grid Supply voltage: 110 V (peak), 1-phase, 50 Hz, SAPF rating: 1.5 kVA, Interfacing Inductor: 3 mH, Non-Linear Load Rating: 1-phase 2-leg Diode Bridge Rectifier connected to $R_{NL1} = 7.5\ \Omega$, $L_{NL1} = 40$ mH, $R_{NL2} = 15\ \Omega$, Reference DC Bus Voltage = 200 V, DC bus Capacitance: 3000 μF, Sampling Time = 50 μs.

Experimental parameters

Grid Supply voltage: 110 V (peak) 1-ϕ, 50 Hz, SAPF rating: 10 kVA, Interfacing Inductor: 3.5 mH, Reference DC Bus Voltage = 200 V, Non-Linear Load Rating: 1-phase 2-leg Diode Bridge Rectifier connected to 250 V, 24 A, Inductive Load : 80 mH, 15 mA, Supply side inductor (L_s):1e−06, DC bus Capacitance: 3000 μF.

References

1. Hooshmand RA, Torabian Esfahani MA (2011) A new combined method in active filter design for power quality improvement in power systems. ISA Trans 50(2):150–158. https://doi.org/10.1016/j.isatra.2010.12.001
2. Li G, Ma F, Wang Y, Weng M, Chen Z, Li X (2020) Design and operation analysis of virtual synchronous compensator. IEEE J Emerg Sel Top Power Electron 8(4):3835–3845. https://doi.org/10.1109/JESTPE.2019.2943723
3. Huang Y, Wang D, Shang L, Zhu G, Tang H, Li Y (2017) Modeling and stability analysis of DC-link voltage control in multi-VSCs with integrated to weak grid. IEEE Trans Energy Convers 32(3):1127–1138. https://doi.org/10.1109/TEC.2017.2700949
4. Li W, Ruan X, Bao C, Pan D, Wang X (2014) Grid synchronization systems of three-phase grid-connected power converters: a complex-vector-filter perspective. IEEE Trans Ind Electron 61(4):1855–1870. https://doi.org/10.1109/TIE.2013.2262762
5. IEEE Power and Energy Society (2014) IEEE recommended practice and requirements for harmonic control in electric power systems. IEEE Std 519–2014(2014):29
6. Eren S, Karimi-Ghartemani M, Bakhshai A (2009) Enhancing the three-phase synchronous reference frame PLL to remove unbalance and harmonic errors. In: 35th annual conference of IEEE industrial electronics, pp 437–441. https://doi.org/10.1109/IECON.2009.5414962
7. Dash A, Behera RK, Bagarty DP, Hota PK (2018) A novel method for synchronization with unbalanced grid: an experimental investigation. In: 2018 20th national power systems conference (NPSC), pp 1–5. https://doi.org/10.1109/NPSC.2018.8771866
8. Zhang C, Zhao X, Wang X, Chai X, Zhang Z, Guo X (2018) A grid synchronization PLL method based on mixed second- and third-order generalized integrator for DC offset elimination and frequency adaptability. IEEE J Emerg Sel Top Power Electron 6(3):1517–1526. https://doi.org/10.1109/JESTPE.2018.2810499
9. Karimi Ghartemani M, Khajehoddin SA, Jain PK, Bakhshai A (2012) Problems of startup and phase jumps in PLL systems. IEEE Trans Power Electron 27(4):1830–1838. https://doi.org/10.1109/TPEL.2011.2169089
10. Xie M, Wen H, Zhu C, Yang Y (2017) DC offset rejection improvement in single-phase SOGI-PLL algorithms: methods review and experimental evaluation. IEEE Access 5:12810–12819. https://doi.org/10.1109/ACCESS.2017.2719721
11. Terriche Y, Mutarraf MU, Mehrzadi M, Lashab A, Guerrero JM, Vasquez JC, Kerdoun D (2019) Adaptive CDSC-based open-loop synchronization technique for dynamic response enhancement of active power filters. IEEE Access 7:96743–96752. https://doi.org/10.1109/ACCESS.2019.2929615
12. Baradarani F, Dadash Zadeh MR, Zamani MA (2015) A phase-angle estimation method for synchronization of grid-connected power-electronic converters. IEEE Trans Power Deliv 30(2):827–835. https://doi.org/10.1109/TPWRD.2014.2362930
13. Lee KJ, Lee JP, Shin D, Yoo DW, Kim HJ (2014) A novel grid synchronization PLL method based on adaptive low-pass notch filter for grid-connected PCS. IEEE Trans Ind Electron 61(1):292–301. https://doi.org/10.1109/TIE.2013.2245622
14. Saxena H, Singh A, Rai JN (2020) Adaptive spline-based PLL for synchronization and power quality improvement in distribution system. IET Gener Transm Distrib 14(7):1311–1319
15. Li X, Zhang H, Shadmand MB, Balog RS (2017) Model predictive control of a voltage-source inverter with seamless transition between islanded and grid-connected operations. IEEE Trans Ind Electron 64(10):7906–7918. https://doi.org/10.1109/TIE.2017.2696459
16. Kumar N, Hussain I, Singh B, Panigrahi BK (2018) Implementation of multilayer fifth-order generalized integrator-based adaptive control for grid-tied solar PV energy conversion system. IEEE Trans Ind Inf 14(7):2857–2868. https://doi.org/10.1109/TII.2017.2777882

17. Bagheri A, Mardaneh M, Rajaei A, Rahideh A (2016) Detection of grid voltage fundamental and harmonic components using Kalman filter and generalized averaging method. IEEE Trans Power Electron 31(2):1064–1073. https://doi.org/10.1109/TPEL.2015.2418271
18. Chittora P, Singh A, Singh M (2017) Performance evaluation of a new Kalman filter based least mean square algorithm for power quality improvement. In: 1st IEEE international conference on power electronics, intelligent control and energy systems, ICPEICES 2016. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/ICPEICES.2016.7853382
19. Golestan S, Guerrero JM, Vasquez JC (2018) Steady-state linear Kalman filter-based PLLs for power applications: a second look. IEEE Trans Ind Electron 65(12):9795–9800. https://doi.org/10.1109/TIE.2018.2823668
20. Karimipour H, Dinavahi V (2015) Extended Kalman filter-based parallel dynamic state estimation. IEEE Trans Smart Grid 6(3):1539–1549. https://doi.org/10.1109/TSG.2014.2387169
21. Panigrahi R, Subudhi B (2017) Performance enhancement of shunt active power filter using a Kalman filter-based H∞ control strategy. IEEE Trans Power Electron 32(4):2622–2630. https://doi.org/10.1109/TPEL.2016.2572142
22. Shah P, Singh B (2020) Adaptive observer based control for rooftop solar PV system. IEEE Trans Power Electron 35(9):9402–9417. https://doi.org/10.1109/TPEL.2019.2898038
23. Modi G, Kumar S, Singh B (2018) Improved Widrow-Hoff based adaptive control of multi-objective PV-DSTATCOM system. In: 2018 5th IEEE Uttar Pradesh section international conference on electrical, electronics and computer engineering, UPCON 2018. Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/UPCON.2018.8597113
24. Karuppanan P, Mahapatra KK (2012) PI and fuzzy logic controllers for shunt active power filter A report. ISA Trans 51(1):163–169. https://doi.org/10.1016/j.isatra.2011.09.004
25. Agarwal RK, Hussain I, Singh B (2018) Application of LMS-based NN structure for power quality enhancement in a distribution network under abnormal conditions. IEEE Trans Neural Netw Learn Syst 29(5):1598–1607. https://doi.org/10.1109/TNNLS.2017.2677961
26. Ahmad MT, Kumar N, Singh B (2016) Fast multilayer perceptron neural network-based control algorithm for shunt compensator in distribution systems. IET Gener Transm Distrib 10(15):3824–3833. https://doi.org/10.1049/iet-gtd.2016.0328
27. Bayindir KÇ, Cuma MU, Tümay M (2006) Hierarchical neuro-fuzzy current control for a shunt active power filter. Neural Comput Appl 15(3–4):223–238. https://doi.org/10.1007/s00521-005-0024-8
28. Arora A, Singh A (2019) Design and implementation of Legendre based neural network controller in grid connected PV systems. IET Renew Power Gener 13(15):2783–2792. https://doi.org/10.1049/iet-rpg.2019.0269
29. Chittora P, Singh A, Singh M (2018) Chebyshev functional expansion based artificial neural network controller for shunt compensation. IEEE Trans Ind Inf 14(9):3792–3800. https://doi.org/10.1109/TII.2018.2793347
30. Ahmad MT, Kumar N, Singh B (2017) AVSF-based control algorithm of DSTATCOM for distribution system. IET Gener Transm Distrib 11(13):3389–3396. https://doi.org/10.1049/iet-gtd.2017.0225
31. Scarpiniti M, Comminiello D, Parisi R, Uncini A (2013) Nonlinear spline adaptive filtering. Signal Process 93(4):772–783. https://doi.org/10.1016/j.sigpro.2012.09.021
32. Scarpiniti M, Comminiello D, Scarano G, Parisi R, Uncini A (2016) Steady-state performance of spline adaptive filters. IEEE Trans Signal Process 64(4):816–828. https://doi.org/10.1109/TSP.2015.2493986
33. Salomon D (2006) Beizer approximation. In: Curves and surfaces for computer graphics, 1st edn. Springer, NY, pp 176–248
34. Racine JS (2018) A primer on regression splines. CRAN R-Project 1990:1–11

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

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

DESIGN AND ANALYSIS OF SINGLE POINT CUTTING TOOL

A.K. Madan[*1], Badal Singh[*2], Devansh Yadav[*3], Deepak Kumar[*4]

[*1]Professor, Mechanical Engineering, Delhi Technological University, Delhi-42, India.

[*2,3,4]B.Tech. Students, Mechanical Engineering, Delhi Technological University, Delhi-42, India.

ABSTRACT

The standard feature method is used to study the effect of different angles on the force applied to the instrument in between to cut. Researchers use this method to better understand the mechanisms of chip formation, heat generation in cutting areas, as well as the features of the tool-chip surface collision in mechanized areas. The value of Vonmisses stress decreases as the rake angle increases. In the present study, mesh was created in the ANSYS once boundary conditions are applied and analysis is performed due to the constraints used. Results are calculated in the software and can be verified.

Keywords: Analysis, Design, Research, Software, Stress.

I. INTRODUCTION

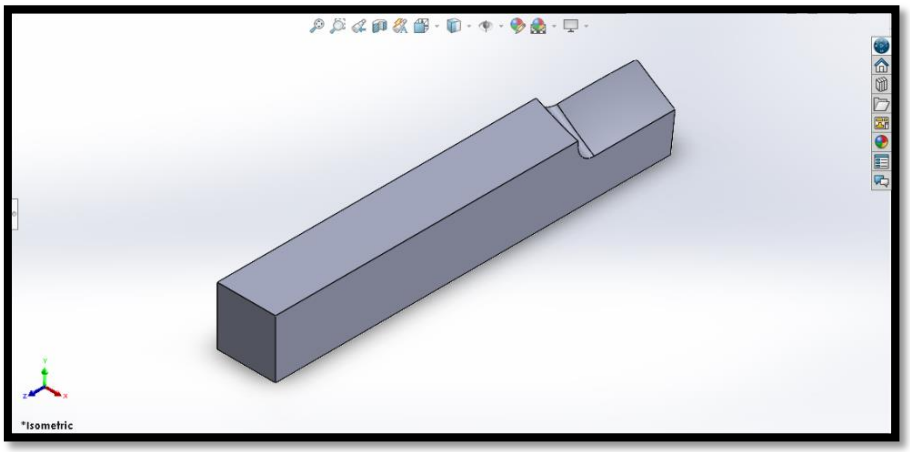
Machining is the process in which a material is cut to a desired final shape and size by a controllable material-removal process. It is the most widespread process for the shaping of metal, it has become a very significant aspect of modern society and industry. Machining is a common fabrication technique in which metal is removed from a part using a tool with a small, hard tip.Over the history of machining, guidelines and conventions have arisen based on empirical information of trade-offs between cutting speed and tool replacement time. Machining is a term covering a large collection of manufacturing processes designed to remove material from a work piece. The purpose of this paper is to provide a review of efforts related to single-point cutting tool and finite element analysis. The review is done to offer insight to how the cutting tool behaves when it is acted upon by various loads.

The use of the software , **ANSYS** has been very helpful for determining the deformation of the cutting tool under the influence of maximum loads.

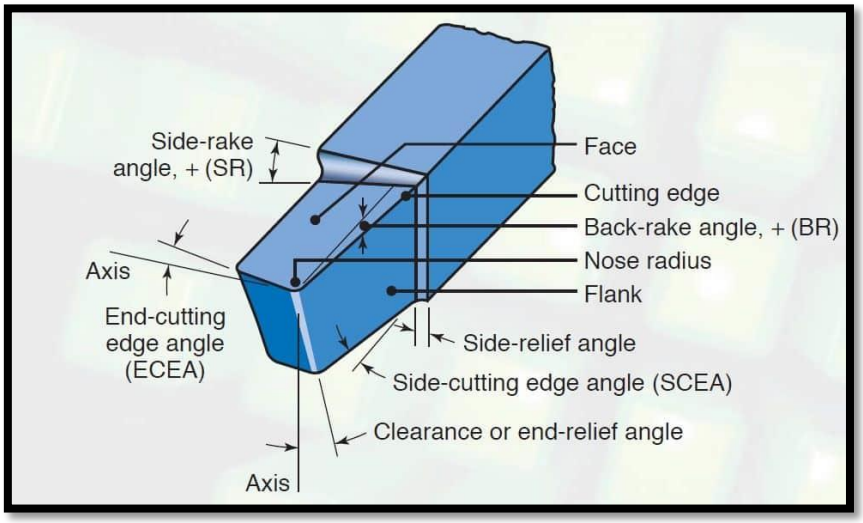
INTRODUCTION OF SINGLE POINT CUTTING TOOL

The single point cutting tool is used for turning, boring, shaping and planning operations, that is,tool used on Lathes, boring machines, shaper, planer etc. are single point cutting tool.

GEOMETERY (SINGLE POINT CUTTING TOOL)



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:04/April-2022 Impact Factor- 6.752 www.irjmets.com



1. Side cutting edge angle:

This angle also is known as the lead angle. This is the angle between the side cutting edge and side of the tool shank.

2. End cutting edge angle:

This is the angle between the end cutting edge and a line normal to the tool shank.

3. Side relief angle:

It is the angle between the portion of the side flank immediately below the side cutting edge and a line perpendicular to the base of the tool and measured at the right angle to the end flank.

4. End relief angle:

It is the angle between the portion of the end flank immediately below the end cutting edge and a line perpendicular to the base of the tool and measured at the right angle to the end flank.

5. Back rack angle :

It is the angle between the tool face and a line parallel to the base of the tool and measured in a plane perpendicular through the side cutting edge.

The back rack angle is positive if the side cutting edge slopes downwards from the point towards the shank and The back rack angle is negative if the slope is side cutting edge is reversed.

6. Side rack angle:

It is the angle between the tool face and a line parallel to the base of the tool and measured in a plane perpendicular to the base and the side cutting edge.This angle gives the slope of the face of the tool from the cutting edge.

II. LITERATURE REVIEW

A review of previous research efforts related to single-point cutting tools and limited feature analysis is presented in it this chapter. There are also reviews of other relevant research studies. Rogério Fernandes Brito, Solidônio Rodrigues de Carvalho, Sandro MetrevelleMarcondes de Lima e Silva, JoãoRoberto Ferreira [1] investigates the effects of heat on cutting tools considering the thickness of the coating and the temperature fluctuations.Diamond tools, K10 substrates, and TiN and Al2O3 coatings were used. Numerical methods are applied usingANSYS CFX software. Numerical analysis involves solid objects with known and consistent boundary conditionsthermophysical properties. The test is used to confirm the proposed operation.

L.B.Abhang and M. Hameedullah [2] analyzed the experimental data to improve the first and second systemmathematical models depending on the parameters of the application using the response area function. Using the appthermocouple technique, EN-31 steel alloy with tungsten carbide inserts was modified. Results analysis is availabledone mathematically and graphically. Metal cutters, speed, feed values, cutting depth, and toolbar radius areproper.

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

Maheshwari N Patil, ShreepadSarange [3] explained how tools and temperatures can be calculated for usenumerical model of metal cutting processes. In the test, it is clear that with increasing depthof cut, temperature at tool tip also increased. Moreover, as the power of the tools increases with the depth of cutting, they also increasedepth of cut. Tool failure is mainly due to this problem. Additionally, we see that the tool vibratesabout 2.5 mm in cut depth. Heat is emitted too much in this case, leading to faint tools. The setting is designed to measure power during dynamometer cutting and tool response analysis.

According to Sambhav et al. [4], the standard definition of piercing used in the grinding parameters was straightforwardCAD must be used as opposed to NURBS. However, a mathematical model that includes standard SPCTto facilitate digestion needs to be improved, as described in this paper.

III. MATERIAL AND METHODOLOGY

HIGH SPEED STEELS

High speed tool includes a set of steel alloys called for their cutting capacity faster than traditional high carbon steel tools previously used for cutting tools. This is due to the unusual strength, resistance to scratches and resistance to softness at high temperatures, due to the composite materials and heat treatment used. We follow the American Society for Testing and Materials in describing high-speed steel as given in Specification A600-79 — high carbon steel containing tungsten and / or molybdenum, as well as chromium, vanadium and sometimes cobalt. High-speed instruments (HSS) are likely to continue to be used in the foreseeable future in many applications such as drilling, reshaping, pressing and dying, forming, brushing and grinding due to their comfort which can be molded with both soft materials. and the situation is difficult.

STEEL 4340

AISI 4340 steel (UNS G43400) is a very strong medium carbon low alloy steel, which combines deep durability, high ductility, durability and strength, and has high wear resistance and crawling resistance. It is particularly susceptible to irritation and embrittlement, and shows good energy retention at high temperatures and softens easily. In small parts, the metal is hardened in the air. By doing so, the oil is often extinguished.

IV. MODELLING AND ANALYSIS

Computer-aid design (CAD) uses pc structures (or workstations) for use during arrival, conversion,testing, or format development. CAD software enhances fashion designer product, improves format,helps to communicate and create production details. CAD output remains in state of electronic documents for printing, manufacturing, or various production activities. CADD time (soComputer Aid Design and Drafting) is also used. Its use in designing visual systems is called digital designautomation, or EDA. Mechanical format automation (MDA) or computer-aid drafting (CAD), i.e.the process of creating technical drawings with the help of computer programs, is what we call a machinedesign. The CAD software design system uses both vector-primarily based entirely on shortcutsshow custom draft gadgets, or they can also produce beautiful images that reflect the normal look of gadgets designed.

4.1. INTRODUCTION TO SOFTWARES USED

INTRODUCTION TO SOLIDWORKS

Solidworks, is a three-dimensional modeling software.engineering, design, production, and editing of CAD. CAD programs that use devices based on parameter law tocapture product behavior. These systems use limits, size and power.. Solidworks says it may provide a more efficient building feeling thandifferent modeling software due to its unique capabilities that integrate parameter integration as welldirect modeling in one place. A complete set of applications includes product development,to provide designers with alternatives for all stages of the process. The software system is also greatan excellent customer interface that gives maximum pleasure to designers. In addition it has interpersonal skillsthat makes it easier to share designs and make changes. There are countless benefits of using Solidwporks. We will test them in this feature series.First, the most important advantage is the extended production due to its green design and flexibility.It is finally designed to be easy to use and has the ability to allow formatting methods to move the briefer, creatingan increase in the degree of production of a fashion designer.Another feature is that the software system is available in 10 languages. PTC knows it has someonepeople from all over the world use their software, to provide it in a few languages as well

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

INTRODUCTION TO ANSYS

Structural Analysis :-

ANSYS Autodyne is a pc simulation tool for simulating the response of building materials in a nutshell.period of unbalanced loading from impact, high difficulty or explosion. ANSYS Mechanical ANSYS Mechanical is aan analytical device for the limited aspect of structural assessment, as well as direct, indirect and dynamic research. Thispc simulation product provides limited features in version behavior, and assists fabric models and mathematical modelsa wide variety of machine design issues. ANSYS Mechanical also contains hot tests as wellsimilar physics skills in acoustics, piezoelectric, thermal-structural and thermo-electric tests.

Fluid Dynamics:-

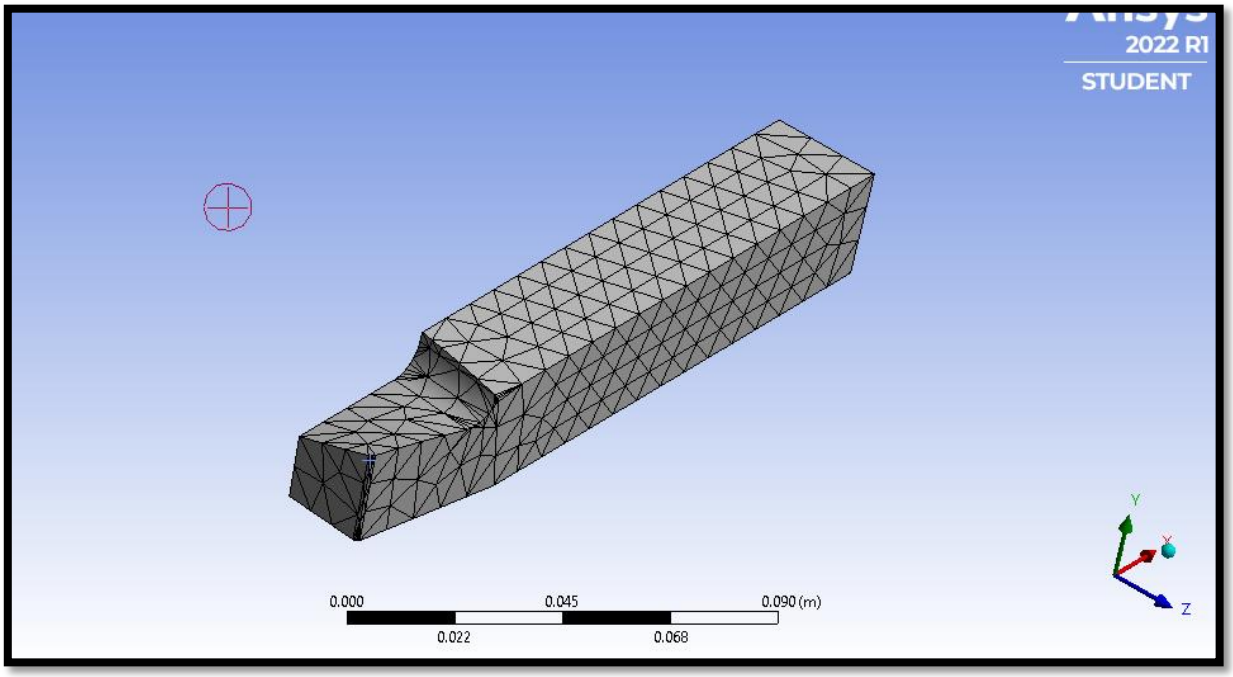
ANSYS Fluent, CFD, CFX, FENSAP-ICE and related software programComputational Fluid Dynamics program material used by engineers for planning and testing. Thisequipment can mimic the flow of fluid in a digital environment - for example, the energy of a liquid in a ship;gas turbine engines (including compressors, fire chamber, mills and afterburners); flightaerodynamics; pumps, lovers, HVAC buildings, mixed vessels, hydro cyclones, vacuum cleaners, and more forward.

Explicit Dynamics :-

Time management method used in the Explicit Dynamics analysis system. It is so named because the way it calculates the answer in the present tense uses clear information Once the body is properly placed, the next step is to define the initial conditions or boundary conditions. At least one first condition is required to complete the setup.After defining the initial conditions (initial velocity, Angular velocity), the analysis setting should be maintained according to the need for the problem. In the case of analysis, time measures should be clearly defined. The timing of the solution depends on the steps of time.

MESHED MODEL

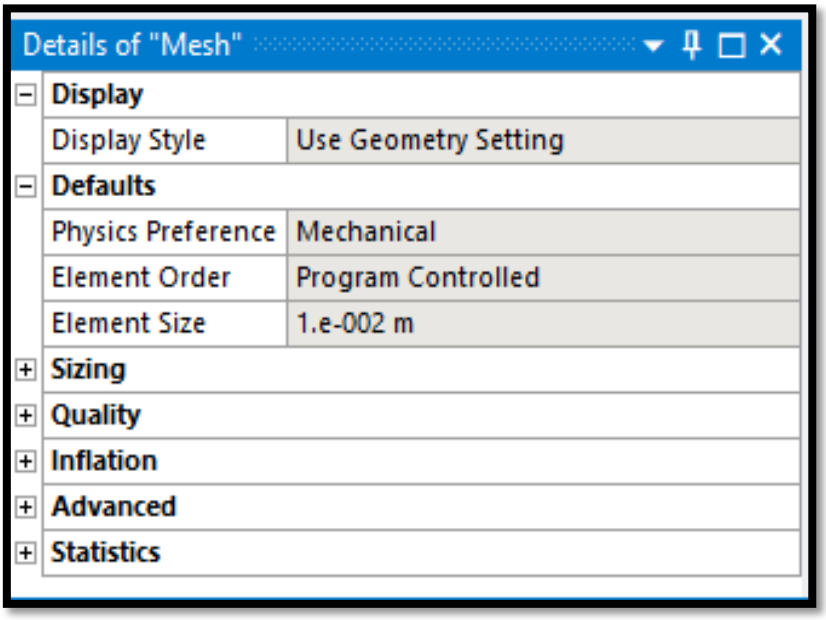
Meshing is one of the most important aspects of getting accurate results from FEA/FEM and CFD simulations.



4.2. DETAILS OF MESH

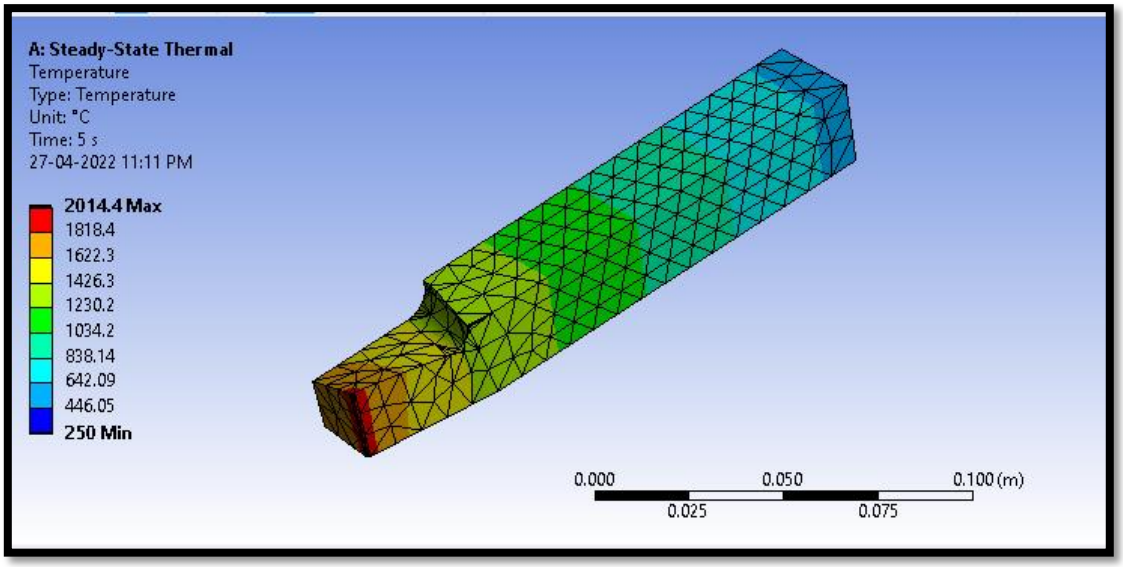
Element size = 0.01

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

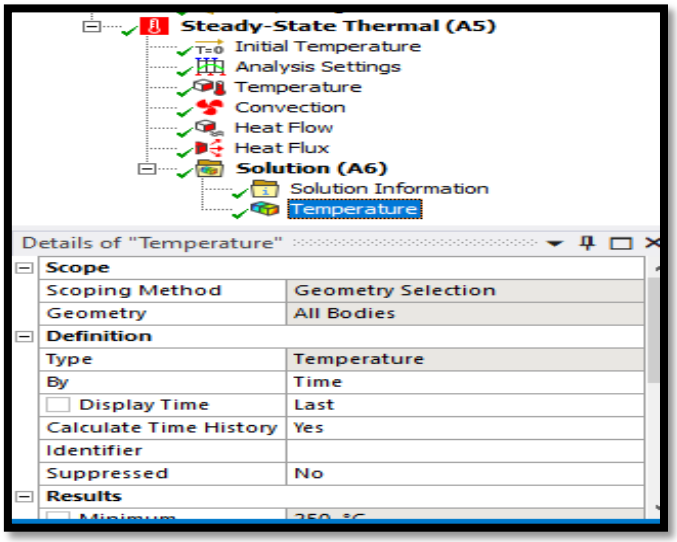


STEADY STATE THERMAL

Solid temperature analysis checks the temperature balance of the system where the temperature remains unchanged over time. In other words, the analysis of the robustness of solidity involves assessing the equilibrium of the system under constant temperature loads and environmental conditions. An easy way to analyze the solid state is to analyze the line position in which input parameters, such as material structures, are determined by independent variables.

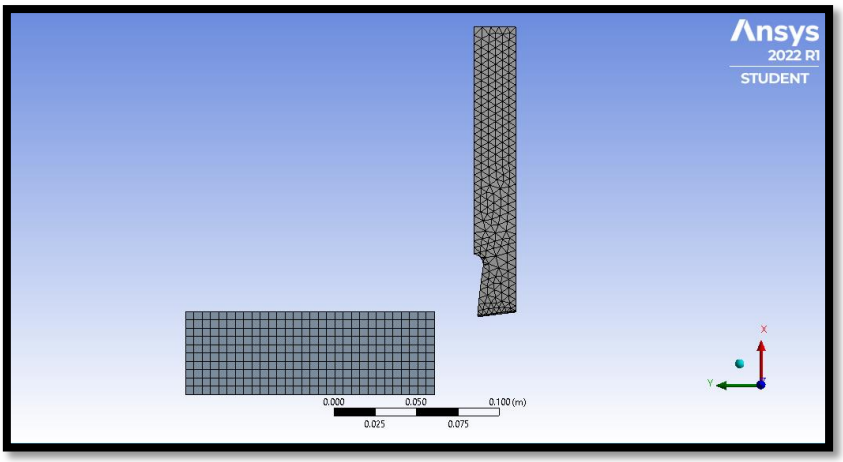


e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:04/April-2022 Impact Factor- 6.752 www.irjmets.com



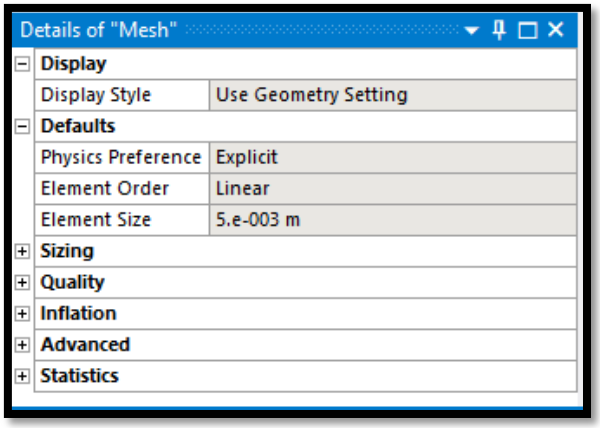
CUTTING OPERATION

Cutting is manufacturing by removal of material. Material particles, chips, are mechanically extracted from raw material or unfinished part by cutting the edges of the tool. The tool has one or more cutting edges, which can be geometrically defined by number, shape, and shape (cutting with cut-out-cut geometric edges), and in the process of damaging chips separation occurs with multiple cutting edges that have to be mathematically defined. and which are distributed randomly within the tool (cutting with geometric cutting edges). Figure shows the cutting processes.



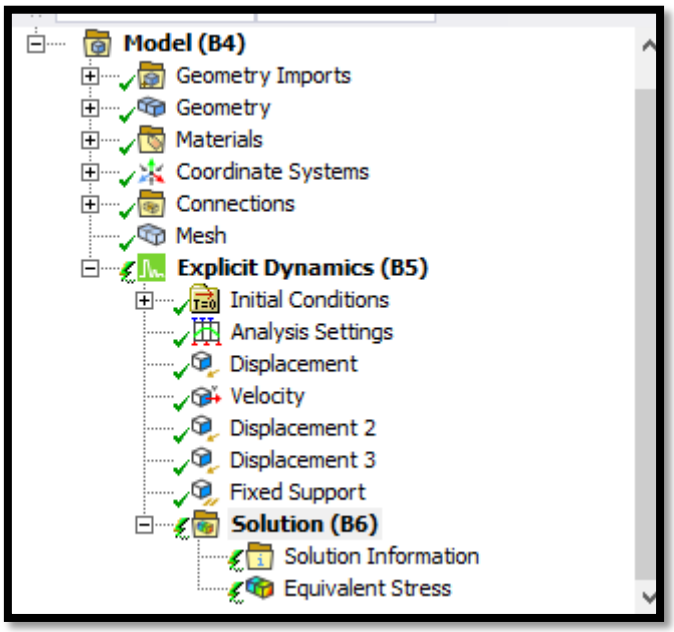
Material of single point cutting tool -> HIGH SPEED STEEL

Material of workpiece -> STEEL 4340



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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

OUTLINE TREE

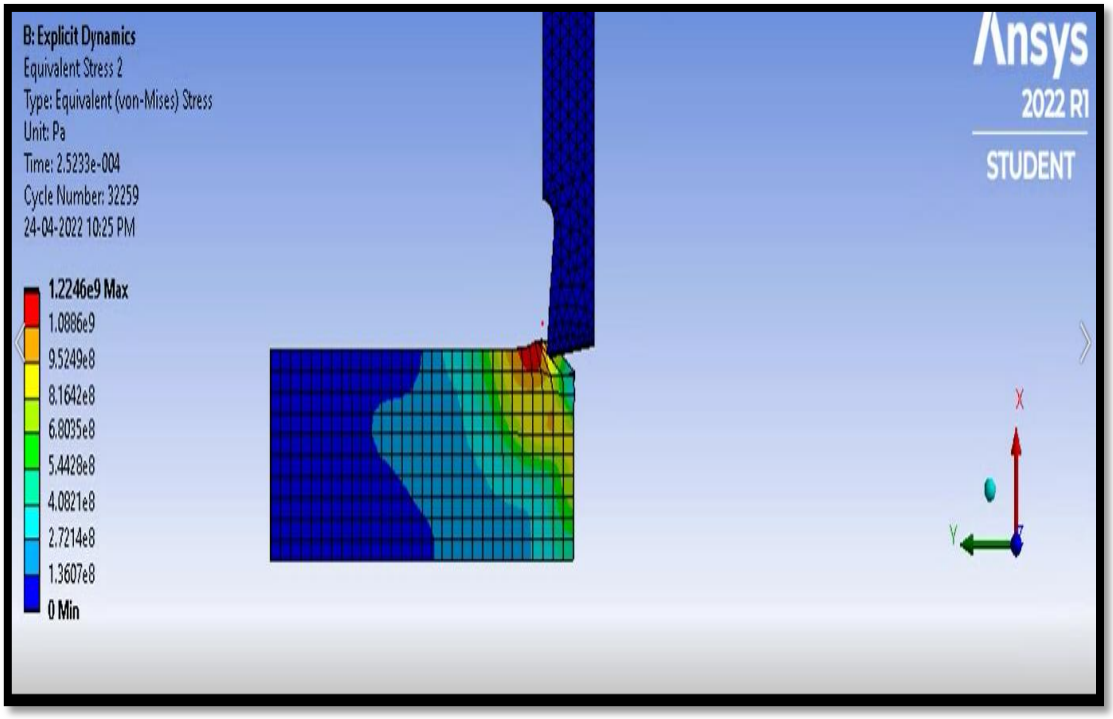


EQUIVALENT STRESS

Equal pressure is actually the result of a shear strain power output per unit measured at different points in the emphasis and helps determine the probability of the failure of the said objects in terms of Von Mises failure. It is often used for ductile substances.

$$\sigma_v = \sqrt{\frac{(\sigma_1 - \sigma_2)^2 + (\sigma_2 - \sigma_3)^2 + (\sigma_3 - \sigma_1)^2)}{2}}$$

STRESS



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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

V. CONCLUSION

In our study the design, optimization and analysis of single point cutting tool was carried out with consideration of parameters like stress, equivalent stress, steady state thermal using various softwares like solidworks and ansys. The analysis part contains various operations and analysis like cutting operation depicted clearly and the thermal analysis part whose animations are given. We get to know how single point cutting tool is made, what are materials and procedures in single point cutting tool and how it works.

VI. REFERENCES

[1] Black, J. T. and Huang, J. M., "An evaluation of chip separation criteria for the fem simulation of machining", ASME Journal of Manufacturing Science and Engineering, 1995.

[2] Marusich, T.D. and Ortiz, M., "Modeling and simulation of high-speed machining," International Journal for Numerical Methods in Engineering1995.

[3] Ozel, T. and Altan, T., "Determination of workpiece flow stress and friction at the chip-tool contact for highspeed cutting," International Journal of Machine Tools and Manufacture, 2000.

[4] Ozel, T. and Zeren, E., "Finite element method simulation of machining of AISI 1045 steel with a round edge cutting tool," Proceedings of the 8th CIRP International Workshop on Modeling of Machining Operations, Chemnitz, Germany, 2005.

[5] Ozel, T., "Modeling of Hard Part Machining: Effect of Insert Edge Preparation for CBN Cutting Tools," Journal of Materials Processing Technology, 2003.

[6] Ozel, T., "The influence of friction model on finite element simulations of machining", International Journal of Machine Tools and Manufacture, 2005.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

DESIGN AND MODELLING OF SQUARE HOLE DRILLING MACHINE

Dr. Ashok Kumar Madan[*1], Chaudhary Jatin[*2], Chiranjeev Singh Sethi[*3], Deepak Kumar[*4]

[*1,2,3,4]Department Of Mechanical Engineering, Delhi Technological University, Delhi, India.

ABSTRACT

This paper describes the production of square holes in the industry. The production of square holes is very common but is hampered by high production costs. The number of machine tools that produce round holes is thick. But in many engineering areas, square and non-circular holes are required. Currently, non-circular holes are manufactured using CNC machines or hackers. But there is a shortage of machine tools to produce square holes and polygons at low cost. So we have made an effort to design and make an attachment for a drilling machine that can produce square holes easier than the current method. The square hole has received many applications in the areas of wood, marble, and metalwork for various purposes. There is no way to make a square hole but to make a square hole with a drilling function is a different concept and here the concept of a reuleaux triangle is used to make a square hole. This will be very helpful to close the bolt heads etc. Attachments can be inserted into a drilling machine or into a straight drill. The complete design of this 40 mm square size attachment has been successfully completed.

Keywords: Reuleaux Triangle, Drilling Machine, EDM, EN36, Square Hole, Etc.

I. INTRODUCTION

Hole serves an important purpose in various parts of the machine. These holes can be circular, square, rectangular or in any other shape depending on the need or design. With circular holes, equipment is easily accessible . But with square holes, a method such as an electrode-discharge (EDM) machine and an electro-chemical machine is currently being used. These are costly and require special tools. The reuleaux triangle is one such invention discovered by German engineer Frank reuleaux. Although he was not the first to consider the shape formed by the joining of three circles in the corner of an equal triangle. However, the use of this Polygonal hole-making curve was introduced by Mr. James Watt in 1914 and geometry has been emerging on a daily basis. The reuleaux triangle is an illustration of a wide range of geometric detection-like Mobious strip that did not find much practical application until it was too late in the evolution of human intelligence. Until about 1875, the famous engineer Frank Reuleaux discussed the Reuleaux triangle, which came to be used in many ways by the Watts Brother Tool works.

II. LITERATURE REVIEW

1) The reuleaux triangle is made up of circular arcs focusing on the vertices of the equilateral triangle of the 's' side. Now that the radius is equal to s and one of the vertices as centre, construct an arc connecting the other vertices. Similarly, draw the arcs that connect the ends of the other two sides. These arcs form the 'Classic Reuleaux triangle' One of its features is that the width is fixed, meaning that the profile can be completely rotated between two equidistant lines separated by a distance of s.

2) A design review on cam geometry to minimize the effect of fillet radius on Square Hole Drilling Operation. In modern life we have to make a square hole in various things like wall, wood etc. Making a square hole is one of the biggest problems. But making a square hole with a drill bit is a different concept from what we learned the concept of the reuleaux triangle type drill cutter used to make a square hole. By using the reuleaux triangle cutting method we can produce a square hole that is not exactly square because there is a radius of fillet left in each corner of the square that needed more work to be exactly square. To prevail over these problems different types of drilling cutters are introduced.

III. METHODOLOGY

The equations below will explain this phenomena correctly. Let us take an equiangular triangle of side s as shown in the figure. In right angle triangle CRA.

CA=s

RA=0.5 X s

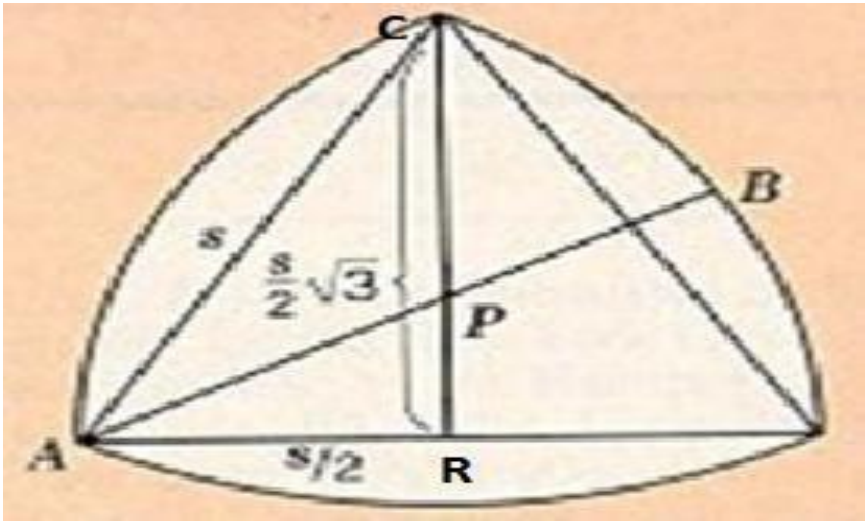
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

RC= 0.866 X s

In triangle ACR

PA= 0.667 X CR= 0.577 X s

Considering reuleaux triangle PB= s-PA= s-0.577s= 0.423s



Reuleaux triangle

The geometric centroid does not stay fastened, nor does it rotate in a circle. It travels a path which consists of four arcs of an ellipse.

IV. PROBLEM STATEMENT

Removal of material from electronic discharge equipment requires the production of waste in the workplace composed of electrode particles and electrical decay products. Contamination of a uniformly distributed gap to a certain limit is desirable in order to close the gap. However, excessive debris is confined to isolated spaces in space due to insufficient flexibility leading to repeated local evacuation. This will have negative effects on process strength, geometry and mechanical stability. Adequate gap cleaning is required for both machine production and quality of machine space. Hitting the water may be accompanied by a forced flow of dielectric fluid into the holes of the tool. But drilling holes leave steps in the machine tool, as the nature of the work produced on the machine corresponds to that of the tool. Cleaning can also be for small holes made especially in the tool. In the event that it is not possible to provide suction holes in any of the electrodes, the dielectric can be controlled from space by means of a plane outside the machine area. This method does not work when the front repair area is large in size. The current work is about and directing the design and implementation of a kinematics novel inspired by the RT concept. Curvilinear rotating tools for dipping regular and irregular polygonal holes with sharp corners, wire EDM is used. If broaching does not work and provides good quality results with local finishing. If the level of stiffness is not allowed to control and the blind take a square hole is required, then this provides a limit to these functions. There should be a tool that can provide the necessary cutting and adhere to production equipment easily and accurately. If the tool is easily available and economical to use, it will be directly beneficial for the manufacturing industries and the consumers.

V. SCOPE

The width of this thesis is to machine a tool with three cutting edges and selecting the appropriate material (EN9) all over the machine, by connecting the tools to the end of the RT. With options instead of tools of various sizes can be machine tool of various sizes. Polygonal holes made by drilling than broaching are superior in a variety of ways.

1. If holes are drilled than broached or press operated, stronger and better parts can be made.
2. Broaching is only beneficial if the required components are large. Drilling is therefore helpful as small items can be produced in a cost-effective and efficient manner.
3. There is no need to pierce the undercuts and the hole has flat unrestricted undercuts.

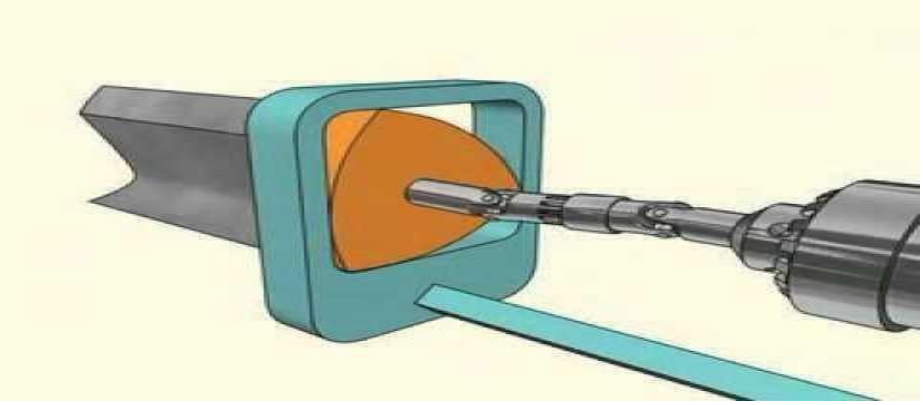
e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:04/Issue:05/May-2022 Impact Factor- 6.752 www.irjmets.com

VI. RESULT EXPECTED

Traditional methods of making square holes are expensive and time-consuming. With the evolution of a square hole drilling machine, Processing of square holes can be done at lesser cost and in a short time. The machine is small in size, which allows flexibility to generate square holes at reduced production costs. This example proves that square holes can be produced using the reuleaux triangle and the universal joint arrangement.

VII. WORKING PRINCIPLE

The main purpose of making a special tool for achieving set goals is to design a method that will convert the shaft movement along the longitudinal axis into a rotating motion on the same axis at a specific profile controlled by four ellipses at Each corner, having their centre at the vertices of the holding square. It will direct the tool to the same profile and keep the rotation tight. The rotation of the tool and the chuck is at same rpm, which is needed to overcome the large amount of force required to cut the metal part. The Revolution becomes an important part so that the centre of the reuleaux triangle is not fixed and must be conveyed to the path created by those four ellipses. After following the basic instructions the need arises to assemble all the parts without neglecting the function of each component.



Square Hole Drilling Machine

REALEUAX TRIANGLE

The side of the equilateral triangle taken to construct the realeuax triangle=S=25mm Length of the arc from each edge= 20mm

Area of the RT= S*S/2(π - √3)= 439.98mm2(approx..)

Purpose of designing a tool:

VIII. TOOL DESIGN

1. Bring down the total cost of building the product by producing acceptable components at lesser cost
2. Enhance productivity by designing components that will make parts faster and sustain quality by manufacturing tools that will consistently produce the parts with the needed accuracy and precision.
3. Lower the cost of special tools by making designs as cost-effective and structured as possible
4. Design tools which will be safe and easy to use

IX. COMPONENTS

1) Point cutting tool-: A special tool mounted on the RT having a 3 cutting edge to obtain the square hole. It is built of EN 36, with high hardness. Predrilling is highly advised, this reduces the amount of tooling and the amount of swarf to be removed from it. It also reduces some of the pressure of the tool and that is why the tool has a great life. The tool is mounted on a reuleaux triangle on the 25mm side, the tool is made of Supporting member-: The square guide is the immobile part that directs the reuleaux triangle to rotate the fixed plane, the square guide is connected to the drill machine using two metal rods, the rods are fastened to the drill machine using clampers. Clampers are provided to attach the supporting members to the portable drilling machine. To get a smooth RT run inside the square guide, the square hole is made wider than the RT, so we can ensure the rotation of the tool without jumping inside the square hole.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

2) Drilling Machine-: In order to provide circular movement to the RT and the tool, impact drill is used .The universal joint end is connected to the drill machine tool holder. The spinning speed does not change for all jobs, while the cutting speed always varies with the cutting edge. The cutting speed is usually calculated by the outer diameter. The cutting speed is zero at the chisel edge center; at any point on the surface and is equal to the radius of that point. This variation in cutting speed at the cutting edges is an important aspect of drilling.

3) Flexible Couplings-: It is necessary to transfer the rotating movement in a square action. This can be accomplished with the help of flexible couplings. Oldham non-coaxial joint, double universal joint and Richard Schmidt integration can be used to achieve the required movement. Flexible couplings can transfer one type of movement to another. The most important point is that during the rotation of the reuleaux polygon, the centre of rotation will not remain in a constant state. So to transmit the torque from the drill to the cam, It is required to use a non-coaxial coupling which can transfer the torque between the non-aligned axis. Oldham Coupling, hook coupling etc are some examples of non-coaxial couplings.

The drilled hole found is 98% square.

X. RESULTS

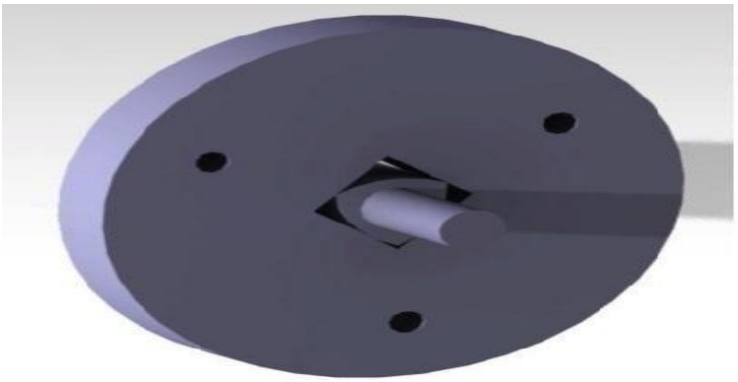
1. The reuleaux triangle system is used successfully to get a square hole in a drill
2. Due to excessive vibration the tool deviates slightly from the previously defined path resulting in an error of 0.2 or 0.3% in the hole drill size.
3. As the base of the machine is not adjusted vibration occurs more frequently

XI. CONCLUSION

After modelling and taking the test on the workpiece, we concluded that drilling a square hole using a drilling machine is possible. The rotatory motion of the drilling machine can be converted into a square movement with the help of flexible coupling. By using this we are able to achieve a 98% square hole. But as we delved deeper into the research we saw that there was a need to make changes to use this concept to be 100% effective as a predrill of a circular shape was needed on a piece of work to become a square later. The base of the machine should be tightly rooted to minimize vibration. Improved results can be obtained by making changes in the design of the tools as the guide support can prevent the movement of the tool as it could not run out of the guide.

XII. FUTURE SCOPE

The concept of square hole drilling machine can be executed on conventional drilling machine. It is only needed to modify a little setup of the machine to provide fixture of the guide plate so that it can reciprocate up and down by the feed mechanism. The following figure shows the plate that can be used for conventional drilling machine.



ACKNOWLEDGEMENT

I am very thankful to Dr AK Madan(Assistant Professor. Mechanical Engineering Dept) and all the faculties of Mechanical Engineering department of Delhi Technological University. They all provided immense support and guidance for the completion of the project undertaken by us. It is with their supervision and guidance that this work came into existence. I would also like to express my gratitude to the university for providing laboratories, infrastructure, test facilities and environment that allowed me to work without any obstructions. I would also like to appreciate the support provided by the lab assistants, seniors and peers who aided me with all the knowledge they had regarding the various topics.

XIII. REFERENCES

[1] International journal for scientific research & development| vol. 4, issue 01, 2016 | issn (online): 2321-0613 Design And Development Of Special Tool To Produce Square Hole Rohit G. Kamble, Manoj Y. Chougule, Rohan

[2] K. Churi, Sanjay R. Lohar

[3] International journal for research in technological studies| vol. 1, issue 5, april 2014 | issn (online): 2348- 1439Review on design of cam geometry for minimization of fillet radius effect in square hole drilling operations R.d.gohil, m.s.kagthara

[4] International journal for engineering applications and technology Modification of drilling tool to make square hole (bangar sunil kisan, prof. Mythili sreeram.) Indian j.sci.res.1(2) : 583-586, 2014

[5] 1corresponding author Square hole created by drilling down by using universal non Shahramhosseini chaleshtori.

[6] Scott G Smith(1998) review on "Drilling Square hole the mathematics teacher" vol 86 pp 579-584

[7] Bangar Sunil Kisan and Prof. Mythili Sreeram 2014 review on "Modification of Square Hole Drilling", IJFEAT journal-ISSN : 2381:8134, pp4-7

[8] S K Hajra Choudhury, A K Hajra Choudhary and Niranjana Roy(2010) review on "Drilling Machines" Elements of Workshop Technology vol 2, pp 227-255[5]

[9] Weisstien, EW. Reuleauxtriangle, Mathworld Mathworld.wolfram.com/reuleaux triangle,html

[10] H .G Eggleston ,1952., A proof of Blaschkle's theorem on Reuleaux Triangle, Quart. J. Math. 3296- 297.

ACADEMIA
Accelerating the world's research.

Design of CDTA and VDTA Based Frequency Agile Filters

Aseem Sayal

Advances in Electronics

Cite this paper

Downloaded from Academia.edu

Get the citation in MLA, APA, or Chicago styles

Related papers

Download a PDF Pack of the best related papers



Realization of CDTA Based Frequency Agile Filter
Aseem Sayal, Richa Choudhary

CDBA Based Universal Inverse Filter
Rajeshwari Pandey

Realization of DVCCTA Based Versatile Modulator
manan tripathi

Hindawi Publishing Corporation
Advances in Electronics
Volume 2014, Article ID 176243, 15 pages
http://dx.doi.org/10.1155/2014/176243

Research Article

Design of CDTA and VDTA Based Frequency Agile Filters

Neeta Pandey,[1] Aseem Sayal,[2] Richa Choudhary,[2] and Rajeshwari Pandey[1]

[1]*Department of Electronics and Communication Engineering, Delhi Technological University, Delhi 110042, India*
[2]*Department of Electrical Engineering, Delhi Technological University, Delhi 110042, India*

Correspondence should be addressed to Neeta Pandey; n66pandey@rediffmail.com

Received 21 May 2014; Revised 3 November 2014; Accepted 3 November 2014; Published 23 December 2014

Academic Editor: Weisheng Zhao

Copyright © 2014 Neeta Pandey 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.

This paper presents frequency agile filters based on current difference transconductance amplifier (CDTA) and voltage difference transconductance amplifier (VDTA). The proposed agile filter configurations employ grounded passive components and hence are suitable for integration. Extensive SPICE simulations using 0.25 μm TSMC CMOS technology model parameters are carried out for functional verification. The proposed configurations are compared in terms of performance parameters such as power dissipation, signal to noise ratio (SNR), and maximum output noise voltage.

1. Introduction

The rapid evolution of wireless services has led to demand for one-fits-all "analog" front end solution. These services use different standards and therefore necessitate development of integrated multistandard transceivers as they result in reduction of size, price, complexity, and power consumption. The parameters of integrated transceiver can be modified in order to be able to adapt to the specifications of each standard [1]. Practically, the designs employ either elements handling various standards in parallel or reconfigurable elements. The frequency agile filter (FAF) [1–10] characterized by adjustment range, reconfigurability, and agility may be used in transceivers. The term shadow filters is sometimes used in literature to refer to FAF [11, 12]. The literature survey shows that a limited number of topologies of active FAF are available and are based on op-amp [1] and current mode active block [2, 3] and CMOS [4].

There is a wide range of current mode building blocks available in open literature. Among these blocks current difference transconductance amplifier (CDTA) [11] is most suitable for current mode signal processing owing to its low input and high output impedances, respectively. The VDTA is yet another recently introduced building block which works on a principle similar to that of CDTA except that the input current differencing unit is replaced by the voltage differencing circuit. Many applications such as filters and oscillators based on CDTA and VDTA are available and have been reported in the literature [13–27] and references cited therein.

The main intention of this paper is to present CDTA and VDTA based frequency agile filter topologies. The proposed filters are suitable for integration as these employ grounded capacitors and a resistor. The paper is organised as follows. The FAF implementation scheme is briefly reviewed in Section 2. The CDTA based Class 0, Class 1, and Class 2 FAF are presented in Section 3. Section 4 deals with the realization of VDTA based Class 0, Class 1, and Class 2 FAF. In Section 5, nonideal analysis of filters is presented. Simulation results are provided in Section 6 to substantiate the proposed FAF topologies. The performance characteristics of filter topologies are described in Section 7. The paper is concluded in Section 8.

2. Implementation Scheme of FAF

The implementation scheme of frequency agile filter (FAF) [3] is briefly reviewed in this section.

2.1. Class 0 FAF. A classical second order filter with band pass (I_{BP}) and low pass (I_{LP}) outputs of Figure 1 is designated as

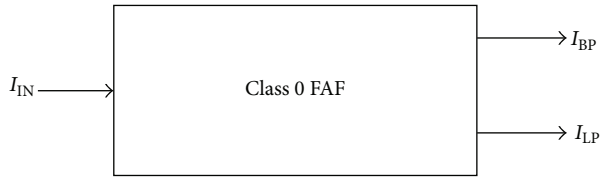


Figure 1: Class 0 FAF [3].

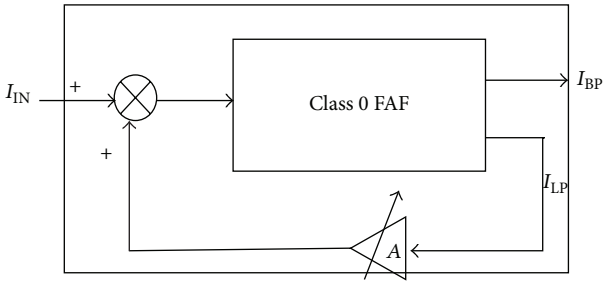


Figure 2: Class 1 FAF.

Class 0 FAF [3]. The transfer functions of Class 0 FAF are given by

$$T_{BP}(s) = \frac{I_{BP}}{I_{IN}} = \frac{ks}{1 + \alpha s + \beta s^2},$$
$$T_{LP}(s) = \frac{I_{LP}}{I_{IN}} = \frac{p}{1 + \alpha s + \beta s^2}. \quad (1)$$

The center frequency (f_0) and quality factor (Q) of the filter are represented by (2) and (3), respectively:

$$f_0 = \frac{1}{2\pi\sqrt{\beta}}, \quad (2)$$

$$Q = \frac{\sqrt{\beta}}{\alpha}. \quad (3)$$

2.2. Class 1 FAF. The basic block diagram of Class 1 FAF is shown in Figure 2 wherein the low pass output of the Class 0 FAF is amplified (with variable gain A) and fed back to the input. The characteristic frequency (f_{0A}) and quality factor (Q_A) of Class 1 FAF are given by (4) and (5), respectively:

$$f_{0A} = f_0\sqrt{(1 + Ap)}, \quad (4)$$

$$Q_A = Q\sqrt{(1 + Ap)}. \quad (5)$$

2.3. Class n FAF. The method outlined for Class 1 FAF realization can be extended for Class n FAF implementation as shown in Figure 3. This requires n amplifiers each with gain A ($A_1 = A_2 = \cdots = A_{n-1} = A_n$) to be placed in n feedback paths obtained in the same way as done in Class 1 implementation. The characteristic parameters of Class n FAF are given by

$$f_{0An} = f_0(1 + Ap)^{n/2},$$
$$Q_{An} = Q(1 + Ap)^{n/2}. \quad (6)$$

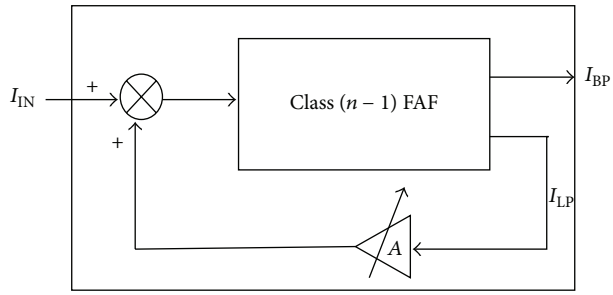


Figure 3: Class n FAF.

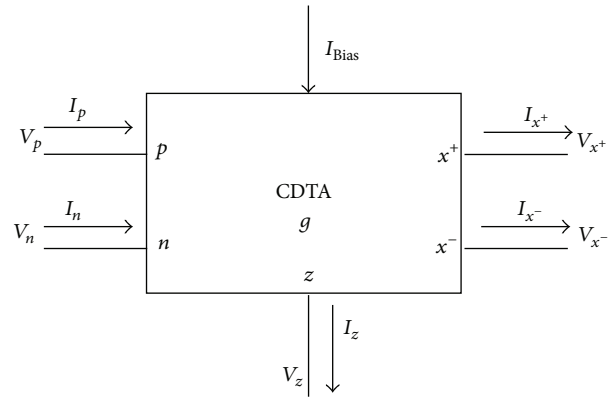


Figure 4: Symbol of CDTA.

3. CDTA Based FAF

The CDTA [11–18] consists of a unity-gain current source controlled by the difference of the input currents and a transconductance amplifier providing electronic tunability through its transconductance gain. The CDTA symbol is shown in Figure 4 and its terminal characteristics in matrix form are given by

$$\begin{bmatrix} V_p \\ V_n \\ I_z \\ I_{x+} \\ I_{x-} \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 & 0 \\ 0 & 0 & g & 0 & 0 \\ 0 & 0 & g & 0 & 0 \end{bmatrix} \begin{bmatrix} I_p \\ I_n \\ V_z \\ V_{x+} \\ V_{x-} \end{bmatrix}, \quad (7)$$

where g is transconductance of the CDTA. The CMOS implementation of CDTA [16] is given in Figure 5. The transistor network comprising transistors Mc1–Mc17 performs [16] current differencing operation on the currents entering at p and n nodes, which is available at Z terminal. The voltage of z terminal drives the source coupled pair (transistors (Mc18–Mc21)) [16] of differential amplifier (Mc18–Mc26) giving a transconductance of g. The value of transconductance (g) is expressed as

$$g = \sqrt{2\mu C_{ox}\left(\frac{W}{L}\right)_{19,21} I_{Bias}}, \quad (8)$$

which can be adjusted by bias current I_{Bias} of CDTA.

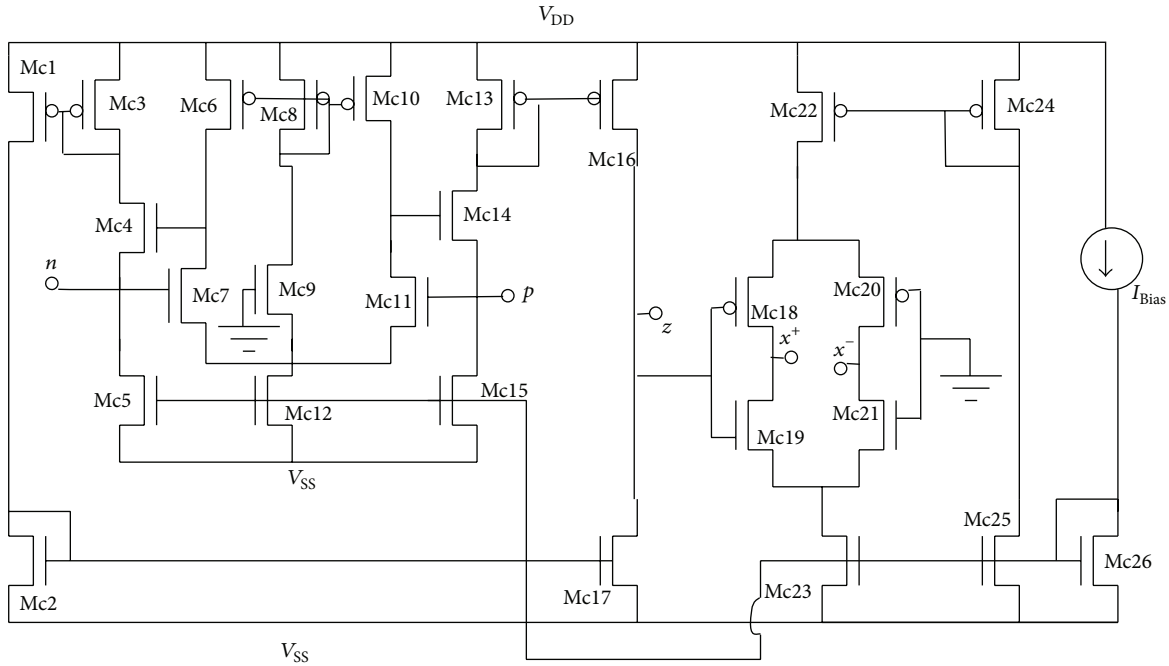


Figure 5: CMOS implementation of CDTA [16].

3.1. CDTA Based Class 0 FAF. The CDTA based second order filter employing two CDTA blocks and two grounded capacitors is shown in Figure 6. The second CDTA block uses additional TA block with its current output terminals denoted by x_c^+ and x_c^-. It provides both low pass and band pass responses at high output impedance and can be used as Class 0 FAF. The current flowing through x^+ and x^- is controlled through transconductance g_2 whereas current flowing through terminals x_c^+ and x_c^- is controlled through g_3. The low pass and band pass transfer functions of CDTA based Class 0 FAF are given by (9) and (10), respectively:

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_2}{C_1 C_2 s^2 + sC_1 g_3 + g_1 g_2}, \tag{9}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{sC_2 g_1}{C_1 C_2 s^2 + sC_1 g_3 + g_1 g_2}. \tag{10}$$

The center frequency and quality factor of Class 0 FAF are expressed as

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{g_1 g_2}{C_1 C_2}}, \tag{11}$$

$$Q = \frac{1}{g_3}\sqrt{\frac{g_1 g_2 C_2}{C_1}}. \tag{12}$$

It may be noted that the Q of the Class 0 FAF can be controlled independent of f_0 by varying g_3.

3.2. CDTA Based Class 1 FAF. The CDTA based Class 1 FAF is shown in Figure 7. It employs Class 0 FAF of Figure 6 along with an additional TA block (provides an output current which is product of its transconductance and voltage difference between noninverting (+) and inverting (−) terminals) and one grounded resistor. The TA block in the feedback path functions as an amplifier with tunable gain A as given in

$$A = g_4 R, \tag{13}$$

where g_4 is the transconductance of TA block and is given by $\sqrt{2\mu C_{ox}(W/L)_{19,21} I_{Bias4}}$.

The low pass and band pass transfer functions of CDTA based Class 1 FAF are given by (14) and (15), respectively:

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_2}{C_1 C_2 s^2 + sC_1 g_3 + g_1 g_2 (1 + Rg_4)}, \tag{14}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{sg_1 C_2}{C_1 C_2 s^2 + sC_1 g_3 + g_1 g_2 (1 + Rg_4)}. \tag{15}$$

The center frequency and quality factor of the CDTA based Class 1 FAF are expressed by (16) and (17), respectively:

$$f_{0A} = \frac{1}{2\pi}\sqrt{\frac{g_1 g_2}{C_1 C_2}}\sqrt{(1 + g_4 R)}, \tag{16}$$

$$Q_A = \frac{1}{g_3}\sqrt{\frac{g_1 g_2 C_2}{C_1}}\sqrt{(1 + g_4 R)}. \tag{17}$$

3.3. CDTA Based Class 2 FAF. The CDTA based Class 2 FAF is shown in Figure 8. It employs two CDTA blocks, two grounded capacitors, three TA blocks, and two grounded resistors. The TA blocks in the feedback path are used as amplifier with tunable gain A. The gain A of TA based

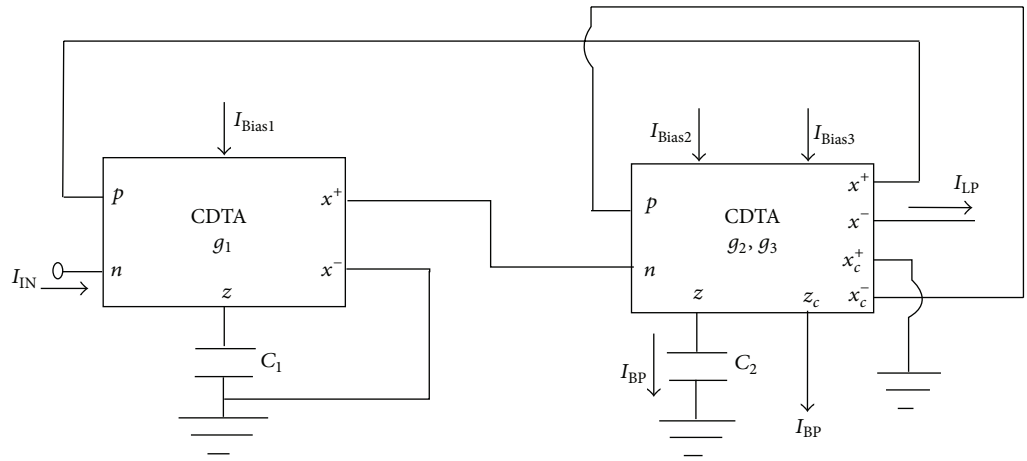


Figure 6: CDTA based Class 0 FAF.

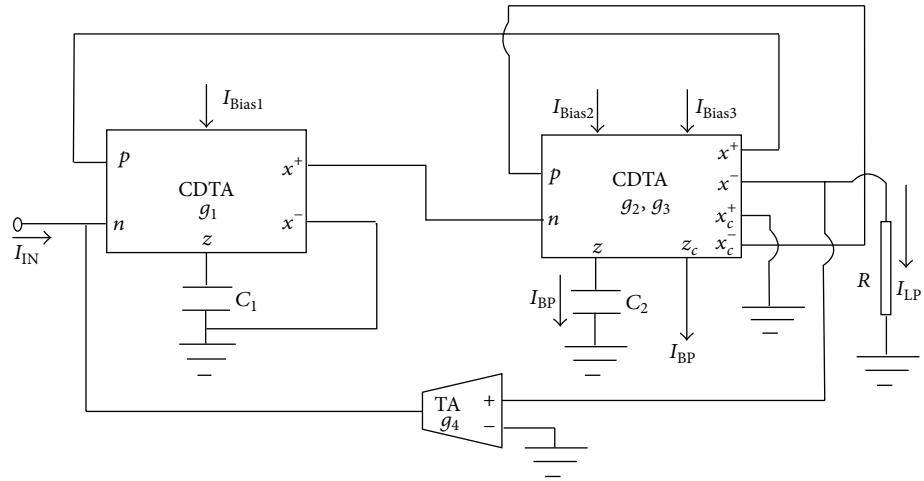


Figure 7: CDTA based Class 1 FAF.

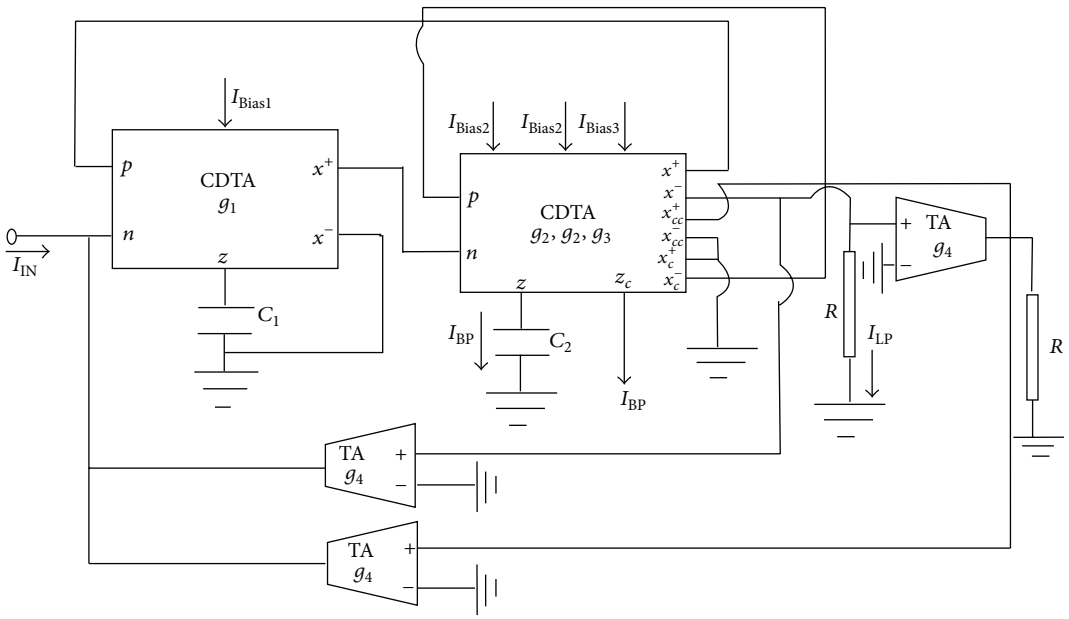


Figure 8: CDTA based Class 2 FAF.

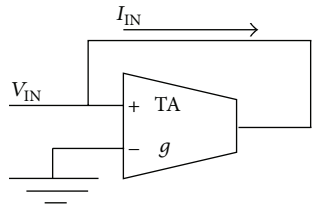


Figure 9: TA realization of a grounded resistor.

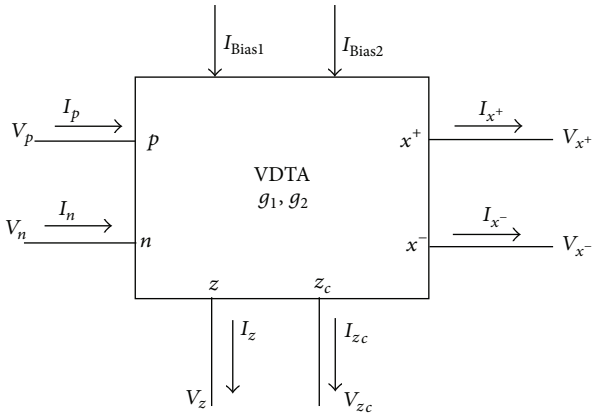


Figure 10: Symbol of VDTA.

amplifier is given by (18). The second CDTA block uses two additional TA blocks. It provides both low pass and band pass responses at high output impedance and can be used as Class 0 FAF. The current flowing through x^+, x^-, x_{cc}^+, and x_{cc}^- is controlled through transconductance g_2 whereas current flowing through terminals x_c^+ and x_c^- is controlled through g_3:

$$A = g_4 R. \tag{18}$$

The low pass and band pass transfer functions of CDTA based Class 1 FAF are given by (19) and (20), respectively:

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_2}{C_1 C_2 s^2 + s C_1 g_3 + g_1 g_2 \left(1 + R g_4\right)^2}, \tag{19}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{s g_1 C_2}{C_1 C_2 s^2 + s C_1 g_3 + g_1 g_2 \left(1 + R g_4\right)^2}. \tag{20}$$

The center frequency and quality factor of the CDTA based Class 1 FAF are expressed by (21) and (22), respectively:

$$f_{0A} = \frac{1}{2\pi}\sqrt{\frac{g_1 g_2}{C_1 C_2}}\left(1 + g_4 R\right), \tag{21}$$

$$Q_A = \frac{1}{g_3}\sqrt{\frac{g_1 g_2 C_2}{C_1}}\left(1 + g_4 R\right). \tag{22}$$

The proposed filter uses grounded resistor of value R (= $1/g$) which can easily be implemented using the TA based structure given in Figure 9.

4. The VDTA Based FAF

The circuit symbol and the CMOS realization of VDTA [20, 21] are shown in Figures 10 and 11, respectively. The VDTA consists of two transconductance (TC) stages termed as input and output stages. The input differential voltage ($V_p - V_n$) is converted to current I_z through TC gain (g_1) of input stage and second stage converts the voltage at z terminal (V_z) to current (I_x) through its TC gain (g_2). The port relations of VDTA can thus be defined by the following matrix:

$$\begin{bmatrix} I_z \\ I_{z_c} \\ I_{x^+} \\ I_{x^-} \end{bmatrix} = \begin{bmatrix} g_1 & -g_1 & 0 \\ -g_1 & g_1 & 0 \\ 0 & 0 & g_2 \\ 0 & 0 & -g_2 \end{bmatrix} \begin{bmatrix} V_p \\ V_n \\ V_z \end{bmatrix}. \tag{23}$$

The TC g_1 and TC g_2 are expressed by (24) which can be adjusted by bias currents I_{Bias1} and I_{Bias2}, respectively:

$$g_1 = \sqrt{2\mu C_{ox}\left(\frac{W}{L}\right)_{1,2} I_{Bias1}},$$
$$g_2 = \sqrt{2\mu C_{ox}\left(\frac{W}{L}\right)_{5,6} I_{Bias2}}. \tag{24}$$

4.1. VDTA Based Class 0 FAF. The VDTA based Class 0 FAF employing single VDTA and two grounded capacitors is shown in Figure 12. This circuit configuration is based on second order filter presented in [21]. However, to allow independent control of quality factor and center frequency an additional TA block with transconductance g_2 is included in VDTA. The current flowing through z terminal is controlled by transconductance g_1 whereas current flowing through z_c terminal is controlled by g_2. The terminal characteristics of the modified VDTA block are given by (25). The low pass and band pass transfer functions of VDTA based Class 0 FAF are given by (26) and (27), respectively:

$$\begin{bmatrix} I_z \\ I_{z_c} \\ I_z' \\ I_{z_c}' \\ I_{x^+} \\ I_{x^-} \end{bmatrix} = \begin{bmatrix} g_1 & -g_1 & 0 \\ -g_1 & g_1 & 0 \\ g_2 & -g_2 & 0 \\ -g_2 & g_2 & 0 \\ 0 & 0 & g_3 \\ 0 & 0 & -g_3 \end{bmatrix} \begin{bmatrix} V_p \\ V_n \\ V_z \end{bmatrix}, \tag{25}$$

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_3}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3}, \tag{26}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{s C_2 g_1}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3}. \tag{27}$$

The center frequency and quality factor of Class 0 FAF are expressed by (28). The center frequency can be controlled by

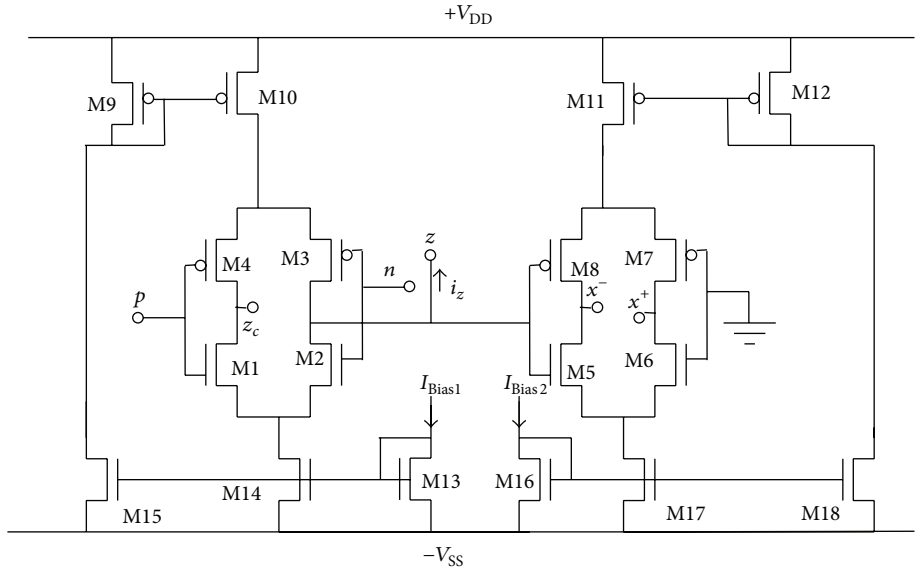


Figure 11: CMOS implementation of VDTA [21].

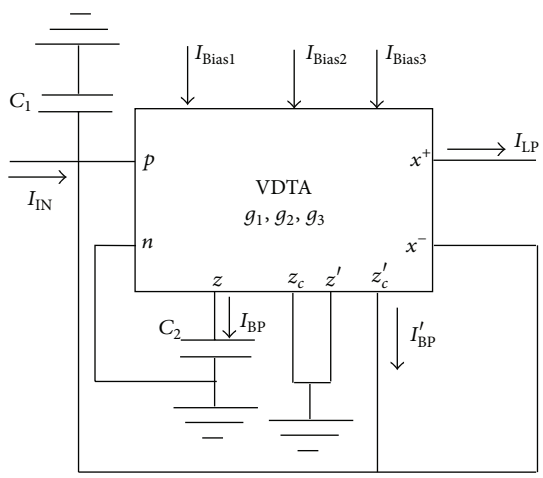


Figure 12: The VDTA based Class 0 FAF.

I_{Bias1} and I_{Bias3} whereas quality factor can be independently controlled by I_{Bias2}:

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{g_1 g_3}{C_1 C_2}}, \qquad Q = \frac{1}{g_2}\sqrt{\frac{g_1 g_3 C_1}{C_2}}. \tag{28}$$

4.2. VDTA Based Class 1 FAF. The VDTA based Class 1 FAF is shown in Figure 13. It employs two VDTA blocks and two grounded capacitors. The second VDTA block is used as amplifier with tunable gain A. The gain A of VDTA based amplifier is given by

$$A = \frac{g_4}{g_3} \tag{29}$$

and can be adjusted by varying I_{Bias3} and I_{Bias4}.

The low pass and band pass transfer functions of VDTA based Class 1 FAF are given by (30) and (31), respectively:

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_3}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3 \left(1 + \left(g_4/g_3\right)\right)}, \tag{30}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{s g_1 C_2}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3 \left(1 + \left(g_4/g_3\right)\right)}. \tag{31}$$

The center frequency and quality factor of the CDTA based Class 1 FAF are expressed by (32) and (33), respectively. The center frequency can be independently controlled by varying I_{Bias2} without changing center frequency:

$$f_{0A} = \frac{1}{2\pi}\sqrt{\frac{g_1 g_3}{C_1 C_2}}\left(\sqrt{1 + \frac{g_4}{g_3}}\right), \tag{32}$$

$$Q_A = \frac{1}{g_2}\sqrt{\frac{g_1 g_3 C_1}{C_2}}\left(\sqrt{1 + \frac{g_4}{g_3}}\right). \tag{33}$$

4.3. VDTA Based Class 2 FAF. The VDTA based Class 2 FAF is shown in Figure 14 which employs three VDTAs, two grounded capacitors and one grounded resistor. The second VDTA block is used as amplifier with tunable gain A. The gain A of VDTA based amplifier is given by (34). The proposed filter uses grounded resistor which can easily be implemented using the TA with transconductance equal to g_3 based structure given in Figure 9. To realize second order filter, I_{Bias7} is set to value of I_{Bias4} such that g_7 is equal to g_4 and I_{Bias6} is set to value such that g_6 is equal to sum of g_3 and g_4; that is, $g_6 = g_3 + g_4$. Consider

$$A = \frac{g_4}{g_3}, \tag{34}$$

which can be adjusted by varying I_{Bias3} and I_{Bias2}, thereby making f_{0A} tunable.

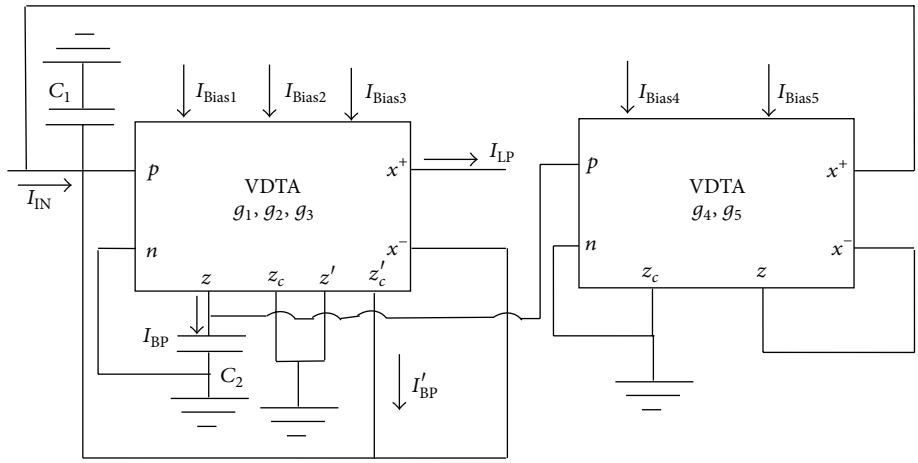


Figure 13: VDTA based Class 1 FAF.

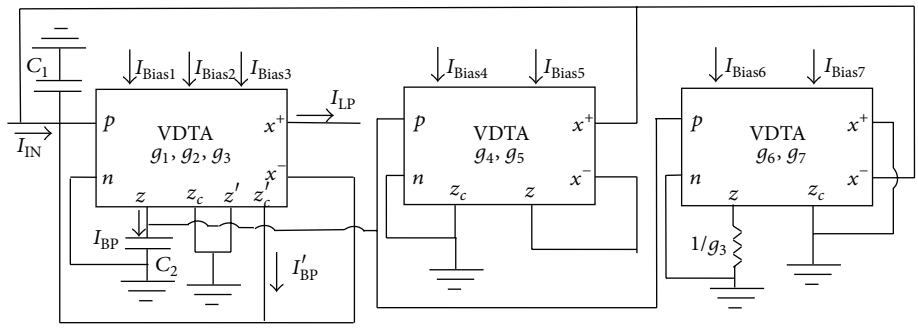


Figure 14: VDTA based Class 2 FAF.

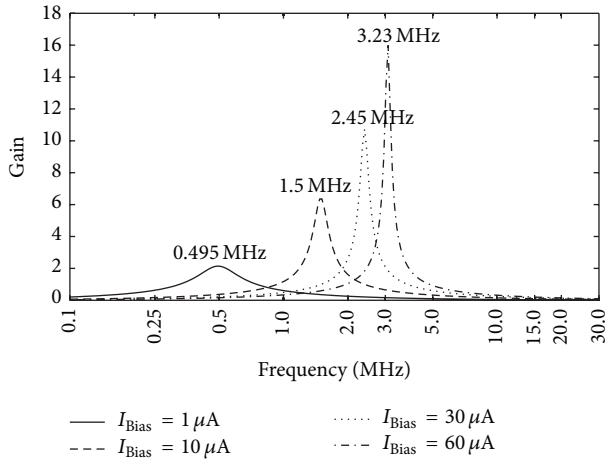


Figure 15: Frequency response of CDTA based Class 0 FAF with $I_{Bias1} = I_{Bias2} = I_{Bias}$.

The low pass and band pass transfer functions of CDTA based Class 1 FAF are given by (35) and (36), respectively:

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_3}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3 \left(1 + (g_4/g_3)\right)^2}, \quad (35)$$

$$\frac{I_{BP}}{I_{IN}} = \frac{s g_1 C_2}{C_1 C_2 s^2 + s C_2 g_2 + g_1 g_3 \left(1 + (g_4/g_3)\right)^2}. \quad (36)$$

The center frequency and quality factor of the CDTA based Class 1 FAF are expressed by (37) and (38), respectively:

$$f_{0A} = \frac{1}{2\pi}\sqrt{\frac{g_1 g_3}{C_1 C_2}}\left(1 + \frac{g_4}{g_3}\right), \quad (37)$$

$$Q_A = \frac{1}{g_2}\sqrt{\frac{g_1 g_3 C_1}{C_2}}\left(1 + \frac{g_4}{g_3}\right). \quad (38)$$

5. Nonideal Analysis

In this section, nonideal analysis of CDTA and VDTA based Class 0 FAF is presented.

5.1. Nonideal Analysis of Class 0 CDTA Based FAF. In practice, the transfer functions (9) and (10) modify due to nonidealities which are classified as tracking errors and parasites. The tracking errors cause current transfer from p and n ports to z port to differ from unity value and are represented by α_p and α_n. There is deviation in transconductance transfer from z to x^+ and x^- ports which is modeled by $\beta g V_z$. The parasites denoted by resistances R_p and R_n are at p and n terminals; shunt output impedances ($R//C$) are present at terminals z, z_c, x^+ and x^-, and x_c^+ and x_c^-. The effect of the parasites is highly dependent on the topology. A close inspection of the circuit of Figure 6 shows that the parasitic capacitances

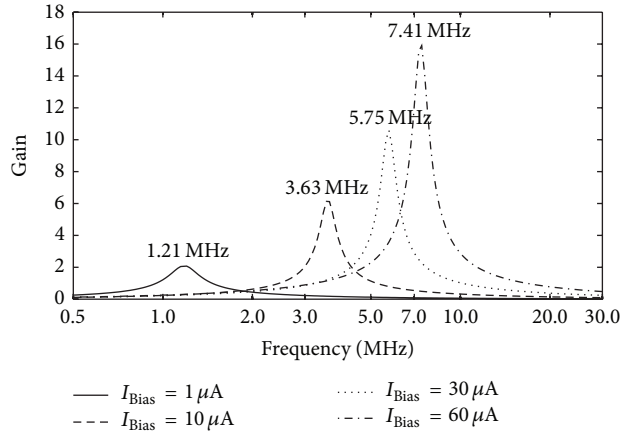


Figure 16: Frequency response of CDTA based Class 1 FAF with $I_{Bias1} = I_{Bias2} = I_{Bias}$.

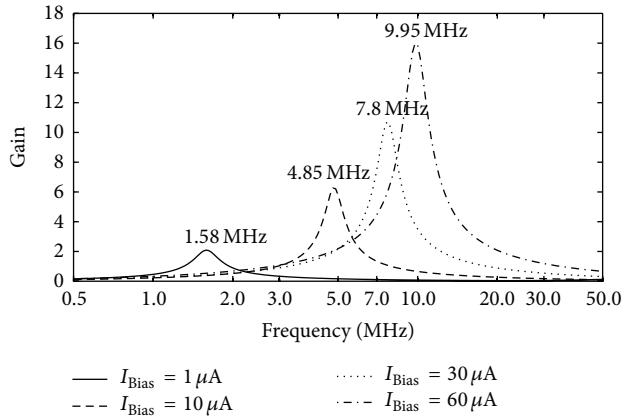


Figure 17: Frequency response of CDTA based Class 2 FAF with $I_{Bias1} = I_{Bias2} = I_{Bias}$.

present at z terminal can be easily accommodated in external capacitances.

Reanalysis of the proposed circuit (Figure 6) yields the following nonideal transfer functions:

$$\frac{I_{LP}}{I_{IN}} = \frac{\alpha_n^2\beta^2 g_1 g_2 Q_1}{D_{n1}(s)}, \tag{39a}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{\alpha_n^2\beta\ g_1\left(sC_{2eq} + G_z\right)Q_1}{D_{n1}(s)}, \tag{39b}$$

where

$$\begin{aligned} D_n(s) &= P_1Q_1\left(sC_{1eq} + G_z\right)\left(sC_{2eq} + G_z\right) \\ &\quad + \alpha_p\beta g_3 P_1\left(sC_{1eq} + G_z\right) + \alpha_p\alpha_n\beta^2 g_1 g_2, \\ P_1 &= \left(1 + G_X R_n + sC_X R_n\right); \\ Q_1 &= \left(1 + G_X R_p + sC_X R_p\right); \\ C_{1eq} &= C_1 + C_z; \qquad C_{2eq} = C_2 + C_z; \\ G_z &= \frac{1}{R_Z}; \qquad G_X = \frac{1}{R_X}. \end{aligned} \tag{39c}$$

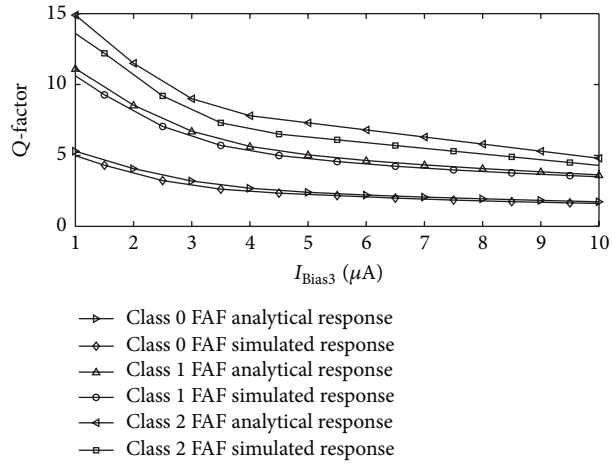


Figure 18: Electronic Q-factor control of CDTA based FAF.

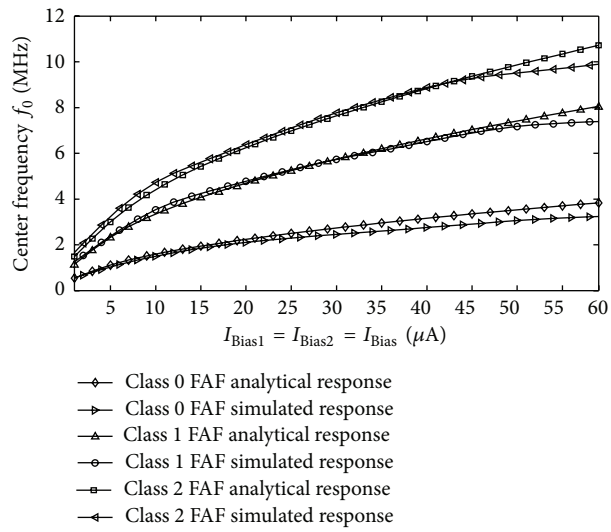


Figure 19: Electronic center frequency control of CDTA based FAF.

Choosing operating frequencies below $\min(1/C_XR_p, 1/C_XR_n)$ (as $G_XR_n \ll 1$ and $G_XR_p \ll 1$) the terms P_1 and Q_1 would not affect the transfer function. For frequencies below $\min(G_Z/C_{1eq}, G_Z/C_{2eq})$, (39c) modifies to

$$D_n(s) = s^2C_{1eq}C_{2eq} + \alpha_p\beta g_3 sC_{1eq} + \alpha_p\alpha_n\beta^2 g_1 g_2, \tag{39d}$$

and transfer functions (39b) and (39c) change to

$$\frac{I_{LP}}{I_{IN}} = \frac{\alpha_n^2\beta^2 g_1 g_2}{D_n(s)}, \tag{40a}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{\alpha_n^2\beta g_1 sC_{2eq}}{D_n(s)}. \tag{40b}$$

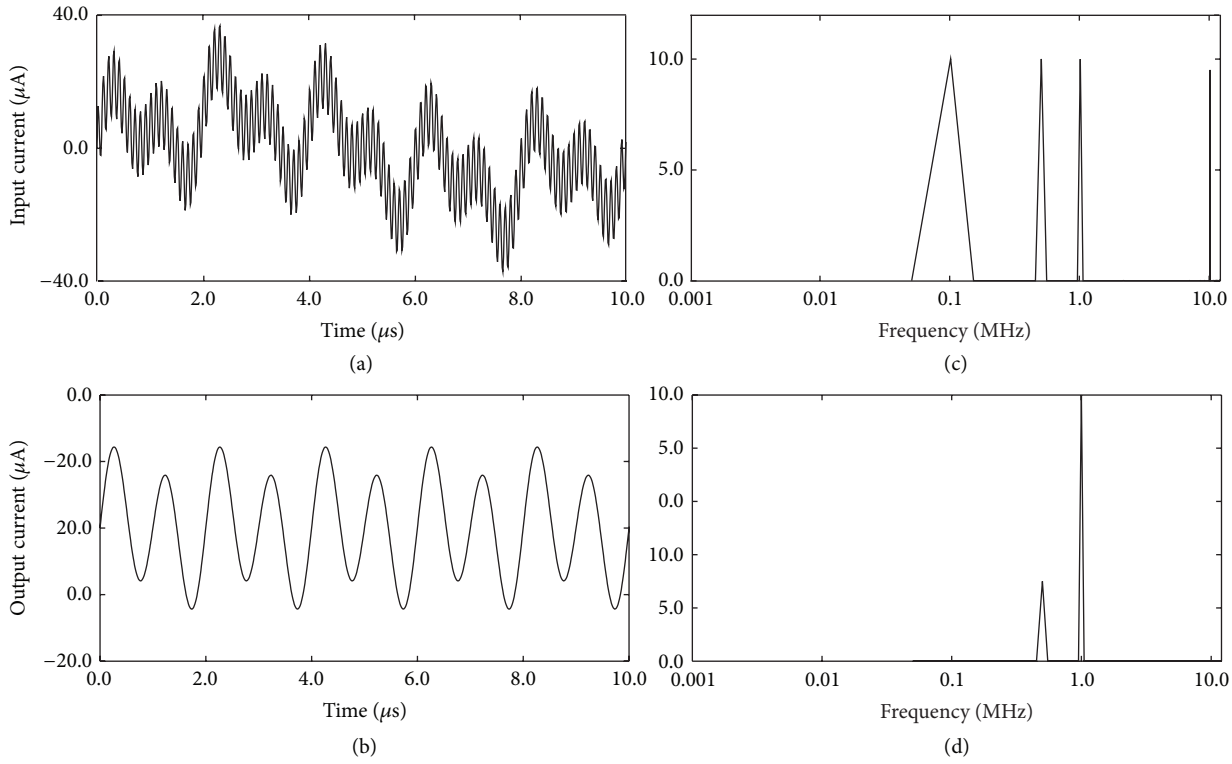


FIGURE 20: ((a) and (c)) Input and its frequency spectrum. ((b) and (d)) Output and its frequency spectrum for Class 0 FAF.

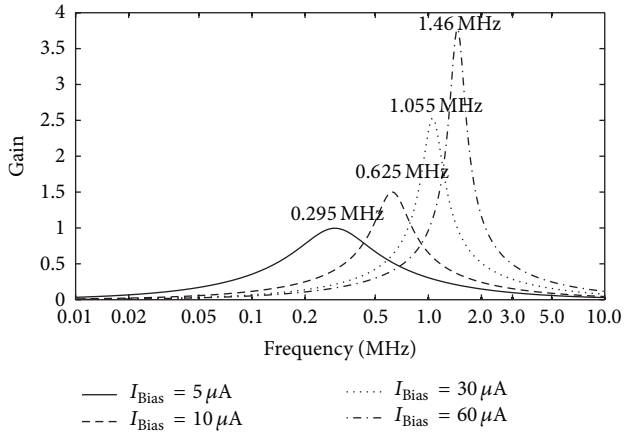


FIGURE 21: Frequency response of VDTA based Class 0 FAF with $I_{Bias1} = I_{Bias3} = I_{Bias}$.

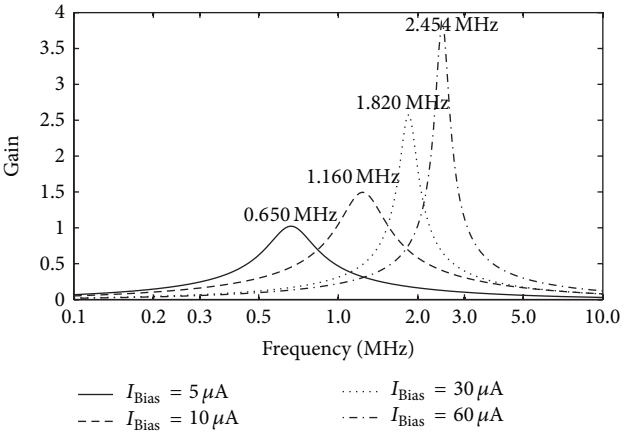


FIGURE 22: Frequency response of VDTA based Class 1 FAF with $I_{Bias1} = I_{Bias3} = I_{Bias}$.

The center frequency, quality factor of Class 0 FAF can be expressed as

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{\alpha_n\alpha_p\beta^2 g_1 g_2}{C_{1eq}C_{2eq}}}, \tag{41a}$$

$$Q = \frac{1}{g_3}\sqrt{\frac{\alpha_n\alpha_p g_1 g_2 C_{2eq}}{C_{1eq}}}. \tag{41b}$$

It is clear that the transfer functions and filter parameters ((40a), (40b) and (41a), (41b)) deviate from the ideal value in presence of nonidealities. The change can, however, be accommodated by adjusting bias currents.

5.2. Nonideal Analysis of Class 0 VDTA Based FAF. Considering the nonideal characteristics of the VDTA, the port relations of current and voltage in (25) can be rewritten as

$$\begin{bmatrix} I_z \\ I_{z_c} \\ I_z^I \\ I_{z_c}^I \\ I_{x^+} \\ I_{x^-} \end{bmatrix} = \begin{bmatrix} \beta g_1 & -\beta g_1 & 0 \\ -\beta g_1 & \beta g_1 & 0 \\ \beta g_2 & -\beta g_2 & 0 \\ -\beta g_2 & \beta g_2 & 0 \\ 0 & 0 & \beta g_3 \\ 0 & 0 & -\beta g_3 \end{bmatrix} \begin{bmatrix} V_p \\ V_n \\ V_z \end{bmatrix}, \tag{42}$$

where β represents the tracking error. Apart from tracking error, the parasites appear as shunt impedances ($R//C$) at

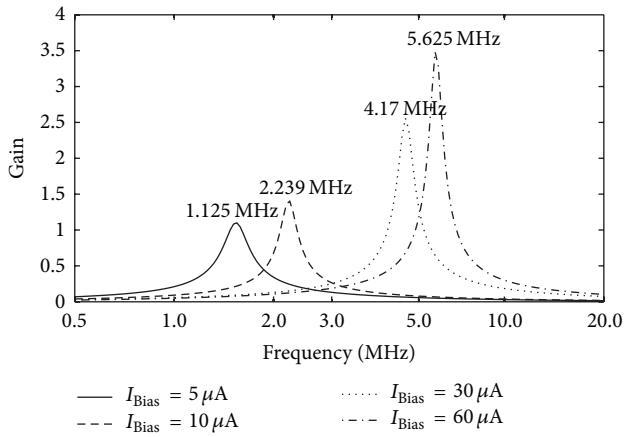


Figure 23: Frequency response of VDTA based Class 2 FAF with $I_{Bias1} = I_{Bias3} = I_{Bias}$.

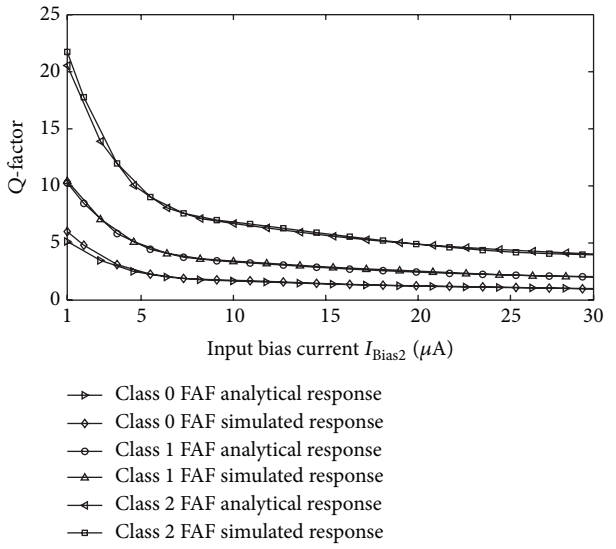


Figure 24: Electronic Q-factor control of VDTA based FAF.

ports p, n, z, z_c, and x^+ denoted by $(R_p//C_p)$, $(R_n//C_n)$, $(R_z//C_z)$, $(R_z//C_z)$, and $(R_x//C_x)$, respectively. The parasitic capacitances present at p, z, z_c, and x terminal can be easily accommodated in external capacitances.

Reanalysis of the proposed circuit in Figure 12 yields the following nonideal transfer functions of Class 0 VDTA based FAF.

Then

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_3 \beta^2}{D_{n2}(s)}, \tag{43a}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{\beta g_1 \left(sC_{2eq} + G_Z\right)}{D_{n2}(s)}, \tag{43b}$$

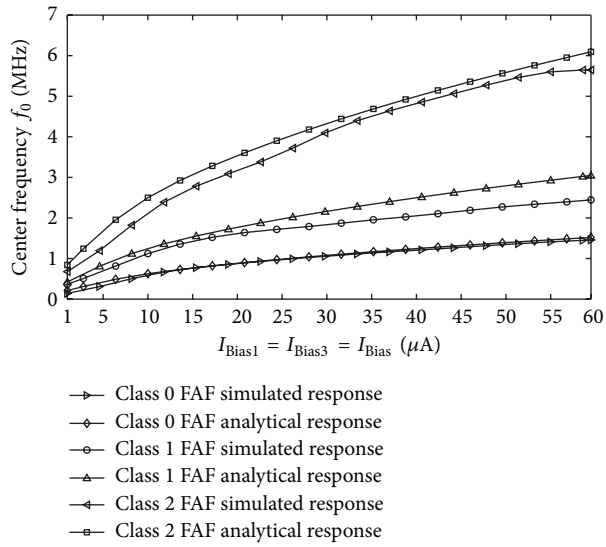


Figure 25: Electronic center frequency control of VDTA based FAF.

where

$$\begin{aligned} D_{n2}(s) = {} & \left(sC_{2eq} + G_z\right)\left(sC_{1eq} + G_x + G_z + G_p + \beta g_2\right) \\ & + \beta^2 g_1 g_3, \\ C_{1eq} = C_1 + C_x + C_z + C_p; \qquad & C_{2eq} = C_2 + C_z; \\ G_z = \frac{1}{R_Z}; \qquad & G_X = \frac{1}{R_X}; \\ G_p = \frac{1}{R_p}. & \end{aligned} \tag{43c}$$

As $G_x + G_z + G_p \ll \beta g_2$, (43c) modifies to

$$D_{n2}(s) = \left(sC_{2eq} + G_Z\right)\left(sC_{1eq} + \beta g_2\right) + \beta^2 g_1 g_3. \tag{44}$$

Choosing operating frequencies below min $(G_Z/C_{1eq}, G_Z/C_{2eq})$ (44) reduces to

$$D_{n2}(s) = s^2 C_{1eq} C_{2eq} + sC_{2eq}\beta g_2 + \beta^2 g_1 g_3, \tag{45}$$

and the transfer function (43b) simplifies to

$$\frac{I_{LP}}{I_{IN}} = \frac{g_1 g_3 \beta^2}{D_{n2}(s)}, \tag{46a}$$

$$\frac{I_{BP}}{I_{IN}} = \frac{\beta g_1 sC_{2eq}}{D_{n2}(s)}. \tag{46b}$$

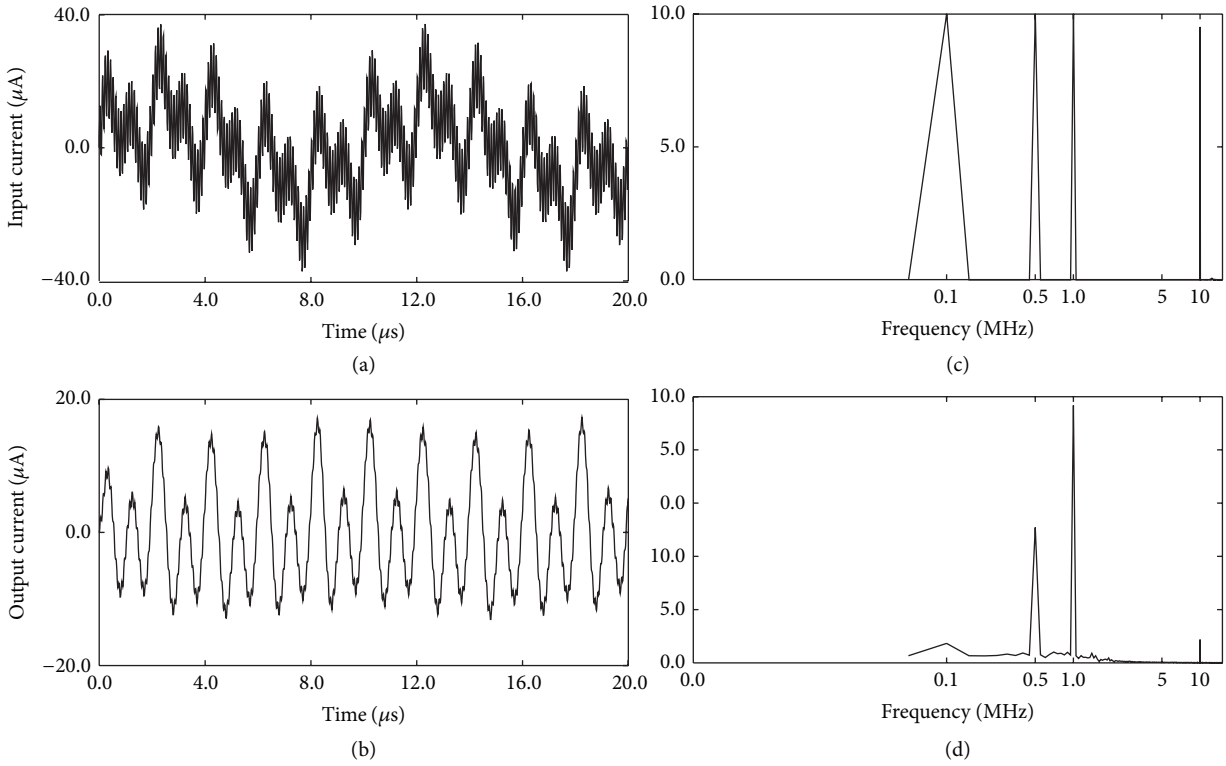


Figure 26: ((a) and (c)) Input and its frequency spectrum. ((b) and (d)) Output and its frequency spectrum for Class 0 FAF.

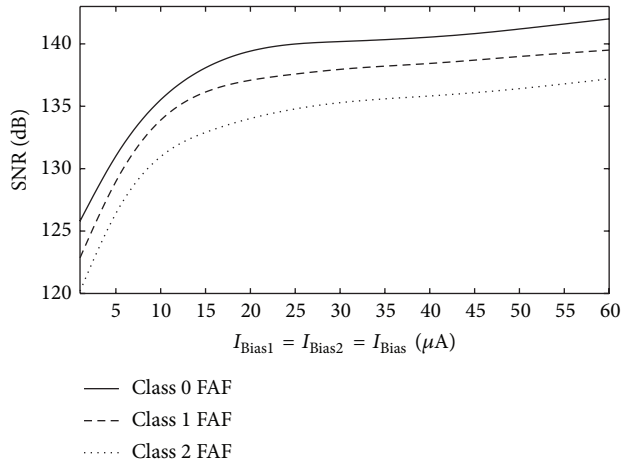


Figure 27: SNR of CDTA based FAF.

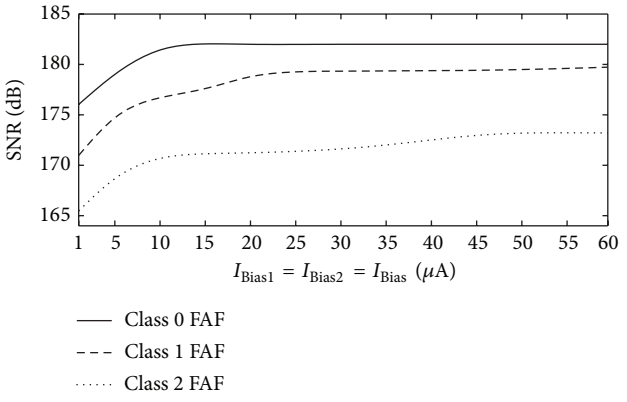


Figure 28: SNR of VDTA based FAF.

And the filter parameters are calculated as

$$f_0 = \frac{1}{2\pi}\sqrt{\frac{\beta^2 g_1 g_3}{C_{1eq}C_{2eq}}}, \tag{47a}$$

$$Q = \frac{1}{g_2}\sqrt{\frac{g_1 g_3 C_{2eq}}{C_{1eq}}}. \tag{47b}$$

It is clear that the transfer functions and filter parameters ((46a), (46b) and (47a), (47b)) deviate from the ideal value in presence of nonidealities. The change can, however, be accommodated by adjusting bias currents.

6. Simulation Results

In this section, the functionality of the proposed filters has been verified. The SPICE simulations results for CDTA and VDTA based filters have been presented using TSMC 0.25 μm CMOS process model parameters and supply voltages of $V_{DD} = -V_{SS} = 1.8$ V.

6.1. Simulation of CDTA Based FAF. The CMOS schematic of Figure 5 is used for verifying CDTA based FAF and the aspect ratios of the MOS transistors are given in Table 1. The additional TA blocks in CDTA providing current ports (x_C^+, x_C^-, x_{cc}^+, and x_{cc}^-) use aspect ratios same as that for x^+ and x^-. The capacitors C_1 and C_2 are chosen as 50 pF each.

Table 1: Aspect ratios of MOS transistors used in CDTA.

MOSFETs	W (μm)/L (μm)
Mc26	7.0/0.7
Mc1	9.8/0.7
Mc2, Mc3, Mc13, Mc16, Mc17	10.5/0.7
Mc19, Mc21	16.1/0.7
Mc6, Mc20	28.0/0.7
Mc8, Mc10, Mc18	28.7/0.7
Mc15, Mc12, Mc5	35.0/0.7
Mc4, Mc14	42.0/0.7
Mc22, Mc23, Mc25	56.0/0.7
Mc24	58.8/0.7
Mc7, Mc9, Mc11	70.0/0.7

The grounded resistor of Figure 7 is realized using TA block. The bias current is set as 0.85 μA to realize a resistor of value 10 kΩ. The frequency responses of CDTA based Class 0, Class 1, and Class 2 FAF topologies are depicted in Figures 15, 16, and 17, respectively. The responses are obtained by varying bias currents I_{Bias1} and I_{Bias2} ($I_{\text{Bias}} = I_{\text{Bias1}} = I_{\text{Bias2}}$) to 1 μA, 10 μA, 30 μA, and 60 μA while keeping I_{Bias3} and I_{Bias4}, respectively, at 0.5 μA and 10 μA. It can be clearly noticed that center frequency f_0 increases on increasing the bias current.

The electronic tunability of quality factor and center frequency for proposed Class 0, Class 1, and Class 2 CDTA based FAF is plotted in Figures 18 and 19, respectively. The analytical and simulated responses describing variation of quality factor are shown in Figure 18 for different values of I_{Bias3} while setting I_{Bias1} and I_{Bias2} to 30 μA. Figure 19 depicts the analytical and simulated responses for center frequency variation for different values of I_{Bias1} and I_{Bias2} ($I_{\text{Bias1}} = I_{\text{Bias2}}$) while setting I_{Bias3} to 0.5 μA. To plot responses for Class 1 and Class 2 CDTA based FAF the bias current I_{Bias4} is taken as 10 μA.

The transient behaviour of proposed CDTA based FAF is also studied by applying input signals of frequencies 100 KHz, 500 KHz, 1 MHz, and 10 MHz, each having an amplitude of 10 μA. The responses for Class 0 FAF are obtained by setting bias currents I_{Bias1} and I_{Bias2} each to 10 μA and I_{Bias3} to 0.5 μA. Figure 20 shows the input and output waveforms along with their frequency spectrum for CDTA based Class 0 FAF. It may clearly be noted that the CDTA based Class 0 FAF allows only 1 MHz signal to pass and significantly attenuates signals of frequencies 100 KHz, 500 KHz, and 10 MHz. Similar responses for Class 1 and Class 2 FAF were also obtained.

6.2. Simulation of VDTA Based FAF. The CMOS schematic of Figure 12 is used for verifying VDTA based FAF and the aspect ratios of the MOS transistors are given in Table 2. The capacitors C_1 and C_2 are taken as 50 pF each. In the realization of Class 2 FAF, the grounded resistor is implemented by TA block. The frequency responses of VDTA based Class 0, Class 1, and Class 2 FAF topologies are shown in Figures 21, 22, and 23, respectively. The responses are obtained by keeping I_{Bias2} to 5 μA and setting bias currents I_{Bias1} and I_{Bias3} ($I_{\text{Bias}} = I_{\text{Bias1}} = I_{\text{Bias3}}$) to 5 μA, 10 μA, 30 μA, and 60 μA. In

Table 2: Aspect ratios of MOS transistors used in VDTA.

MOSFETs	W (μm)/L (μm)
M1, M2, M5, M6	16.1/0.7
M3, M4, M7, M8	28/0.7
M9–M12	21/0.7
M13–M16	7/0.7
M14, M15, M17, M18	8.5/0.7

realization of Class 1 FAF, I_{Bias4} is set to obtain $g_4/g_3 = 3$ while keeping I_{Bias5} equal to I_{Bias1}. In realization of Class 2 FAF, I_{Bias6} is selected such that $g_6 = g_3 + g_4$ while I_{Bias7} is equal to I_{Bias4}.

The electronic tunability of quality factor and center frequency for proposed Class 0, Class 1, and Class 2 VDTA based FAF is plotted in Figures 24 and 25, respectively. The analytical and simulated responses describing variation of quality factor are shown in Figure 24 for different values of I_{Bias2} while setting I_{Bias1} and I_{Bias3} to 30 μA in Class 0. To plot responses for Class 1, I_{Bias4} is set to get $g_4/g_3 = 3$ and I_{Bias5} is set equal to I_{Bias1}. Class 2 FAF responses are plotted by selecting I_{Bias6} to obtain $g_6 = g_3 + g_4$ while I_{Bias7} is equal to I_{Bias4}. Figure 25 depicts the analytical and simulated responses for center frequency variation for different values of I_{Bias1} and I_{Bias3} ($I_{\text{Bias1}} = I_{\text{Bias3}}$) and keeping I_{Bias2} to 5 μA. To plot responses for Class 1, I_{Bias4} is set to a value in such a manner that $g_4/g_3 = 3$ whereas I_{Bias5} is set equal to I_{Bias1}. The Class 2 FAF responses are plotted by setting I_{Bias6} to value such that $g_6 = g_3 + g_4$ whereas I_{Bias7} is equal to I_{Bias4}.

The transient behaviour of proposed agile filter is also studied by applying input signals of frequencies 100 KHz, 500 KHz, 1 MHz, and 10 MHz, each having an amplitude of 10 μA. The responses for Class 0 FAF are obtained by setting bias currents I_{Bias1}, I_{Bias2}, and I_{Bias3} each to 30 μA. The input and output waveforms along with their frequency spectrum for VDTA based Class 0 FAF are shown in Figure 26. It may clearly be noted that the VDTA based Class 0 FAF allows only 1 MHz signal to pass, partially attenuates signal of frequency 500 KHz, and significantly attenuates signals of frequencies 100 KHz and 10 MHz. Similar responses for Class 1 and Class 2 FAF are also obtained.

7. Performance Evaluation

The performance of proposed CDTA and VDTA based FAF circuits is studied in terms of power dissipation, output noise voltage, and SNR. The overall performance characteristics are summarized in Table 3. Figures 27 and 28 depict the signal to noise ratio (SNR) for the proposed CDTA and VDTA based filter topologies for Class 0, Class 1, and Class 2, respectively. The VDTA based FAF proved to be optimum concerning the power dissipation and signal to noise ratio. The maximum output noise voltage is better in VDTA based FAF.

8. Conclusion

In this paper CDTA and VDTA based frequency agile filters are presented. The proposed FAF configurations employ

Table 3: Performance characteristics of CDTA and VDTA based Class 0, Class 1, and Class 2 FAF.

Performance characteristics	Type of FAF	$I_{Bias} = 1\,\mu A$			$I_{Bias} = 10\,\mu A$			$I_{Bias} = 30\,\mu A$			$I_{Bias} = 60\,\mu A$		
		Class 0	Class 1	Class 2	Class 0	Class 1	Class 2	Class 0	Class 1	Class 2	Class 0	Class 1	Class 2
Power dissipation (mW)	CDTA	0.359	0.997	1.63	3.34	3.99	4.63	9.98	10.7	11.3	19.9	20.6	21.2
	VDTA	0.089	0.177	0.405	0.32	1.19	3.45	0.835	3.34	9.57	1.60	6.28	17.4
SNR (dB)	CDTA	124.9	122.1	119.2	135.5	134.2	131.3	140.2	137.9	135.3	142.1	139.5	137.2
	VDTA	175.82	170.4	165.0	181.7	180.5	170.96	182.0	179.3	171.6	182.0	179.75	173.2
Max. output noise voltage (nV)	CDTA	79.37	74.92	75.49	155.02	82.99	58.97	210.46	70.69	40.85	290.59	60.56	31.92
	VDTA	28.5	28.85	28.1	38.92	37.43	51.3	46.10	44.4	38.1	51.62	52.63	25.6

grounded passive components and are suitable for integration. The filter configurations are designed in such a way that quality factor can be independently controlled without changing the center frequency. The simulation results are included to demonstrate the workability of the circuits. The performance of the proposed FAF is evaluated in terms of power dissipation, SNR, and noise performance. The VDTA based FAF proved to be optimum concerning the power dissipation and signal to noise ratio.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

References

[1] P. I. Mak, U. Seng-Pan, and R. P. Martins, *Analog-Baseband Architecture and Circuits for Multistandard and Low Voltage Wireless Transceivers*, Analog Integrated Circuits and Signal Processing, 2007.

[2] Y. Lakys and A. Fabre, "Multistandard transceivers: state of the art and a new versatile implementation for fully active frequency agile filters," *Analog Integrated Circuits and Signal Processing*, vol. 74, no. 1, pp. 63–78, 2013.

[3] Y. Lakys and A. Fabre, "A fully active frequency agile filter for multistandard transceivers," in *Proceedings of the International Conference on Applied Electronics (AE '11)*, pp. 1–7, September 2011.

[4] S. Kaehlert, D. Bormann, T. D. Werth, M.-D. Wei, L. Liao, and S. Heinen, "Design of frequency agile filters in RF frontend circuits," in *Proceedings of the IEEE Radio and Wireless Symposium (RWS '12)*, pp. 13–16, Santa Clara, Calif, USA, January 2012.

[5] A. J. X. Chen, Y. Wu, J. Hodiak, and P. K. L. Yu, "Frequency agile digitally tunable microwave photonic filter," in *Proceedings of the International Topical Meeting on Microwave Photonics (MWP '03)*, pp. 89–92, September 2003.

[6] G. Subramanyam, F. W. van Keuls, and F. A. Miranda, "A K-band-frequency agile microstrip bandpass filter using a thin-film hts/ferroelectric/dielectric multilayer configuration," *IEEE Transactions on Microwave Theory and Techniques*, vol. 48, no. 4, pp. 525–530, 2000.

[7] H. Chandrahalim, S. A. Bhave, R. G. Polcawich, J. Pulskamp, and R. Kaul, "A $Pb(Zr_{0.55}Ti_{0.45})\,O_3$-transduced fully differential mechanically coupled frequency agile filter," *IEEE Electron Device Letters*, vol. 30, no. 12, pp. 1296–1298, 2009.

[8] M. W. Wyville, R. C. Smiley, and J. S. Wight, "Frequency agile RF filter for interference attenuation," in *Proceedings of the 6th IEEE Radio and Wireless Week (RWW '12)*, pp. 399–402, Santa Clara, Calif, USA, January 2012.

[9] H. H. Sigmarsson, J. Lee, D. Peroulis, and W. J. Chappell, "Reconfigurable-order bandpass filter for frequency agile systems," in *Proceedings of the IEEE MTT-S International Microwave Symposium (MTT '10)*, pp. 1756–1759, Anaheim, Calif, USA, May 2010.

[10] Y. Lakys, B. Godara, and A. Fabre, "Cognitive and encrypted communications, part 2: a new approach to active frequency-agile filters and validation results for an agile bandpass topology in SiGe-BiCMOS," in *Proceedings of the 6th International Conference on Electrical and Electronics Engineering (ELECO '09)*, pp. II16–II29, November 2009.

[11] V. Biolkova and D. Biolek, "Shadow filters for orthogonal modification of characteristic frequency and bandwidth," *Electronics Letters*, vol. 46, no. 12, pp. 830–831, 2010.

[12] Y. Lakys and A. Fabre, "Shadow filters—new family of second-order filters," *Electronics Letters*, vol. 46, no. 4, pp. 276–277, 2010.

[13] D. Biolek, "CDTA-building block for current-mode analog signal processing," in *Proceedings of the European Conference on Circuit Theory and Design (ECCTD '03)*, vol. III, pp. 397–400, Krakow, Poland, 2003.

[14] D. Biolek, V. Biolkova, and Z. Kolka, "Current-mode biquad employing single CDTA," *Indian Journal of Pure and Applied Physics*, vol. 47, no. 7, pp. 535–537, 2009.

[15] M. Kumngern, P. Phatsornsiri, and K. Dejhan, "Four inputs and one output current-mode multifunction filter using CDTAs and all-grounded passive components," in *Proceedings of the 10th International Conference on ICT and Knowledge Engineering, ICT and Knowledge Engineering*, pp. 59–62, Bangkok, Thailand, November 2012.

[16] S. K. Pandey, A. P. Singh, M. Kumar, S. Dubey, and P. Tyagi, "A current mode second order filter using dual output CDTA," *International Journal of Computer Science & Communication Networks*, pp. 210–213, 2012.

[17] F. Kacar and H. Kuntman, "A new cmos current differencing transconductance amplifier (CDTA) and its biquad filter application," in *Proceedings of the IEEE (EUROCON '09)*, pp. 189–196, St Petersburg, Russia, May 2009.

[18] D. Biolek, E. Hancioglu, and A. Ü. Keskin, "High-performance current differencing transconductance amplifier and its application in precision current-mode rectification," *International Journal of Electronics and Communications*, vol. 62, no. 2, pp. 92–96, 2008.

[19] A. Uygur, H. Kuntman, and A. Zeki, "Multi-input multi-output CDTA-based KHN filter," in *Proceedings of the IEEE International Microwave Symposium Digest*, pp. 1756–1759, 2010.

[20] W. Chiu, S. I. Liu, H. W. Tsao, and J. J. Chen, "CMOS differential difference current conveyor and their applications," *IEE Proceedings-Circuits Devices Systems*, vol. 143, no. 2, pp. 91–96, 1996.

[21] Z. Wang, "2-MOSFET transresistor with extremely low distortion for output reaching supply voltages," *Electronics Letters*, vol. 26, no. 13, pp. 951–952, 1990.

[22] D. Biolek, M. Shaktour, V. Biolkova, and Z. Kolka, "Current-input current-output universal biquad employing two bulk-driven VDTAs," in *Proceedings of the 4th International Congress on Ultra Modern Telecommunications and Control Systems (ICUMT '12)*, pp. 484–489, St. Petersburg, Russia, October 2012.

[23] J. Satansupa and W. Tangsrirata, "Single VDTA based current mode electronically tunable multifunction filter," in *Proceedings of the 4th International Science, Social Science, Engineering and Energy Conference (ISEEC '12)*, pp. 1–8, 2012.

[24] P. Phatsornsiri, P. Lamun, M. Kumngern, and U. Torteanchai, "Current-mode third-order quadrature oscillator using VDTAs and grounded capacitors," in *Proceedings of the 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE '14)*, pp. 1–4, IEEE, Chiang Rai, Thailand, March 2014.

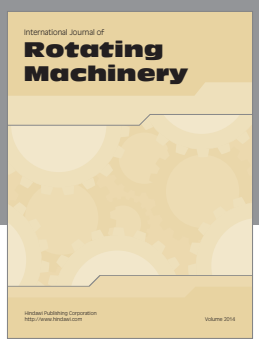
[25] T. Pourak, P. Suwanjan, W. Jaikla, and S. Maneewan, "Simple quadrature sinusoidal oscillator with orthogonal control using single active element," in *Proceedings of the International Conference on Signal Processing and Integrated Networks (SPIN '14)*, pp. 1–4, 2014.

[26] M. Srivastava, D. Prasad, and D. R. Bhaskar, "New Parallel R-L impedance using single VDTA and its high pass filter application," in *Proceedings of the International Conference on Signal Processing and Integrated Networks (SPIN '14)*, pp. 535–537, 2014.

[27] K. Chumwangwapee, W. Jaikla, W. Sunthonkanokpong, W. Jaikhang, S. Maneewan, and B. Sreewirote, "High input impedance mixed-mode biquad filter with orthogonal tune of natural frequency and quality factor," in *Proceedings of the 4th Joint International Conference on Information and Communication Technology, Electronic and Electrical Engineering (JICTEE '14)*, pp. 1–4, Chiang Rai, Thailand, March 2014.

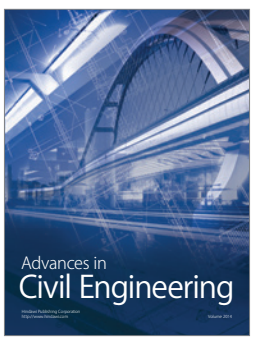


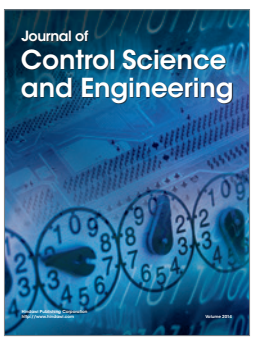






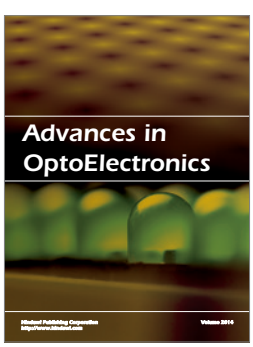






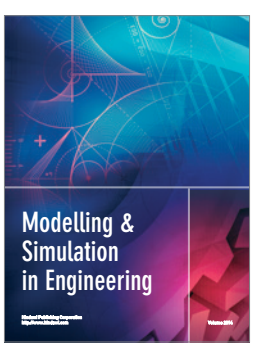


Hindawi
Submit your manuscripts at
http://www.hindawi.com

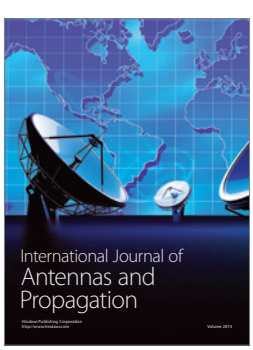




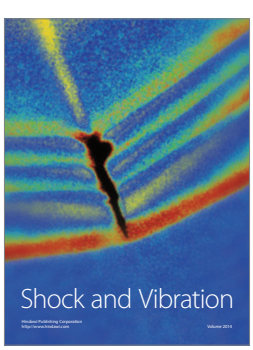


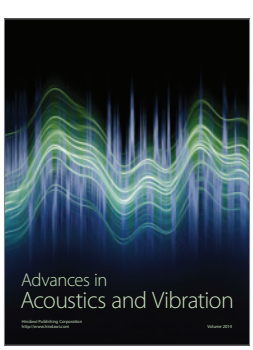












J Mater Sci: Mater Electron

Detection of biomolecules in dielectric modulated double metal below ferroelectric layer FET with improved sensitivity

Yash Pathak[1], Bansi Dhar Malhotra[2], and Rishu Chaujar[1,*]

[1] *Department of Applied Physics, Delhi Technological University, New Delhi 110042, India*
[2] *Department of Biotechnology, Delhi Technological University, New Delhi 110042, India*

Received: 31 December 2021
Accepted: 16 April 2022

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

ABSTRACT

In this work, we examined the double metal below ferroelectric layer FET that is double metal below negative capacitance field-effect transistor (DM-below-NCFET) for biosensing application and change in nanocavity gap with biomolecules as protein, ChO_x (cholesterol oxidase), streptavidin, and uricase. For measuring the electrical characteristic and neutral biosensing such as threshold voltage, switching ratio (I_{on}/I_{off}) of the device which is higher than one without molecules by 1.52 times, sensitivity of protein enhanced by 1.11 over without biomolecule, limit of detection of protein is higher by 1.012 times over without molecule, shift in potential have been researched for cavity length 10 nm. The biosensor indicated improved sensitivity for biomolecules with the rise in their dielectric parameter. Moreover, modulation of the length of the gap of cavity was too examined, exposing that its increment (from 8 to 12 nm) altogether upgraded the sensitivity of the proposed biosensor. Visual TCAD (Technology Computer-Aided Design) software is used for simulating all results. In general, the consequences of this examination represent that such DM-below-NCFET biosensors can display extreme sensitivity (1.11) at small drain voltage (0.4 V), empowering their utilization for biosensor applications to analyze different infections which involve low power, extreme density, and enhanced speed.

1 Introduction

The new COVID-19 epidemic has once more gained a bleeding edge on the impact of microelectronic biosensors in recognition of biomolecules and biological fighting specialists. The applications might be discovered by electronic biosensors uninterruptedly observing inflight biomolecules in conveyance contexts like airplanes and metro rail and midway cooled structures like hospitals, schools, clinics, and so forth. Consequently, the requirement of great importance is to plan and foster quick and precise biosensors. Biosensors utilize chemical responses to distinguish the biochemical mixtures like antibodies,

Address correspondence to E-mail: chaujar.rishu@dtu.ac.in

https://doi.org/10.1007/s10854-022-08290-x
Published online: 07 May 2022

enzymes, nucleic acids, proteins, and so forth. So they are broadly utilized in numerous applications like checking of infections, food investigation, crime detection, ecological field observing, and for the investigation of biomolecules cooperation [1, 2].

Various strategies have been created for recognizing of biomolecules such that the chemical connected immunosorbent measure [3], for Alzheimer's illness, coronary vein infection, and ovarian cancer. In any case, numerous of these methodologies are convoluted and tedious caused by the necessity for modeling procedures [4]. Field-effect transistor (FET) constructed biosensors have drawn in much consideration from late years inferable from their great versatility, extreme sensitivity, quick electrical identification, small force utilization, straight electrical display, and minimal expense large-scale manufacturing in correlation with different strategies, for example, surface plasmon resonance [5], microcantilevers [5], and varieties of fluorescence devices [6, 7]. Minimal expense, profoundly touchy, dependable, easy to use, and speedy symptomatic biosensing gadgets are fundamental for various organic and biomedical applications [8]. For biosensing functions, the main component of key for a MOSFET is sensitivity [9].

Ferroelectric FETs (Fe-FETs) have pulled in improved consideration because of their guarantee in both memory and exchanging applications [10]. Size and power consumption are the main figures of value of any advanced electronic device and the objective of past and present research endeavors is to devise forcefully scaled devices that consume ultralow power [11]. The effect of channel length scaling was excluded by NCFET. Additionally, the last utilizes the Landau model just in the outside-of-plane polarization part, which is not reasonable for huge ferroelectric thicknesses [12]. In our previous works, we examined and analyzed the execution of the double NCFETs with lengthy channels, utilizing logical displaying [13]. The Landau's parameters for the FE material is $\alpha = -1.0 \times 10^{+7}$ and $\beta = 8.9 \times 10^{+8}$ [14]. Double metal below negative capacitance FET (DM-below-NCFET) for protein biomolecule gives a higher ON-state flow contrasted to other biomolecule; be that as it may, we discover this not to be the case for all the ferroelectric materials. Moreover, a definite actual understanding of the conduct of the dual constructions and general execution analysis is yet disappeared in the writing [15–20].

In this article, double metal below negative capacitance FETs (DM-below-NCFETs) for protein biomolecule enhanced the performance of the device in comparison with uricase, streptavidin, ChO_x (Cholesterol oxidase), without biomolecules such as higher sensitivity, lower leakage current, low input power consumption, and control over gate channel device. A new concept of a ferroelectric layer with a dielectric constant (K = 33.5) has been introduced. The matrix of device performance has been extracted by using visual TCAD (Technology Computer-Aided Design) software. In this DM-below-NCFET, the I_{off} is lying in the request for 10^{-12} A and thus has less releasing capacity from a battery point of view, which is useful for superior performance. The DM-below-NCFET has upgraded the sum of I_{on} and great depletion channel at off-state (i.e., at V_{gs} = 0). From this examination, we become more acquainted with the device that has phenomenal execution of analog and power amplification at the point when having an analogy and electrical stimulation between various biomolecules.

A proposed device has been utilized as a biosensor for detecting different biomolecules. The main use of the device is to comprehend different biomolecules, which is done when biomolecule ties to the gate of the device and modifies its surface charge. In the proposed DM-below-NCFET device (biosensor), the device's channel limit gets improved with the electrical curve of biomolecules from which they can be distinguished. In the traditional MOSFET, the current through channel flow when the gate voltage transcends the threshold voltage (V_{TH}) [2, 21, 22]. With a general high doping concentration, the DM-below-NCFET acts as a resistor. At the point when the quantity of charged analyte changes on the sensor surface, the conductance of the resistor changes likewise. Following recently talked about detecting instruments, the charges from analytes influence the conductivity of the device channel due to attractive or repulsion electrostatic force. All in all, the level of depletion inside channel changes because of charges from analytes in the arrangement [23].

2 Structure of device and method of simulation

Figure 1a reveals the structure of Double Metal below ferroelectric layer FET that is double metal below negative capacitance FET and Fig. 1b represent the spacer has biomolecules. For immobilization of biomolecules in the area of Bio1 and Bio2, the region between area source and gate (L_{bio1}) or drain and gate (L_{bio2}) is the same as cavity gap. Since the determining region is the most important characteristic and behavior of DM-below-NCFET, rhe channel's length (L_c) is 20 nm, body's length (L_b) is 60 nm, drain's/-source's length ($L_{d/s}$) is 10 nm, the Bio1/Bio2 length ($L_{bio1/bio2}$) is 10 nm, body's thickness (T_b) is 50 nm, drain's/source's thickness ($T_{d/s}$) is 3 nm, oxide's thickness (T_o) is 1 nm, Ferroelectric's thickness (T_f) is 1 nm, Metal1's (T_{m1}) thickness is 2nm, insulator's thickness (T_i) is 2.3 nm, Metal2's thickness (T_{m2}) is 1.7 nm, the thickness of Bio1/Bio2 is 2 nm, and the thickness of gate is 2 nm. The total gate thickness (T_{gt}) is the addition of thickness oxide, Metal2, insulator, Metal1, ferroelectric layer, and gate. The material of the body is Si (silicon), the material of drain/source/substrate/gate is Al (aluminum), the material of Insulator is nitrite, the material of oxide is SiO_2 (silicon dioxide), the material of Metal1 is Au and Material of ferroelectric is HfO_{2FE} (an innovative idea of a ferroelectric layer with hafnium dioxide). A proposed device DM-below-NCFET contains metal, insulator, metal, ferroelectric, intermediate gate, silicon dioxide, and silicon from top to bottom. The doping concentration of body (N_b) is 1×10^{16} cm^{-3} with uniform profile and acceptor type, the doping concentrating of drain/source ($N_{d/s}$) is 1×10^{20} cm^{-3} with Gaussian profile and type of donor and the doping concentration of threshold voltage (N_{vt}) is 1×10^{18} cm^{-3} with Gaussian profile and acceptor type. The work function is very high kept at 5.4 eV for a Metal1 gate.

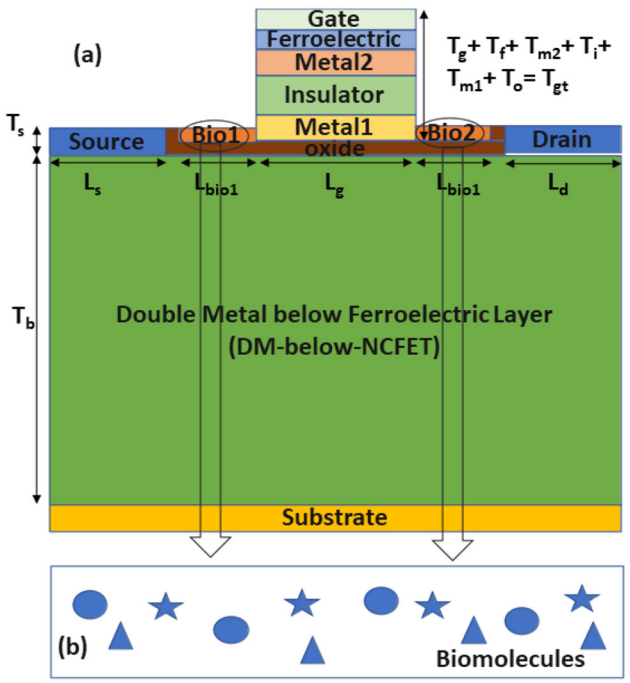


Fig. 1 Schematic diagram of double metal below ferroelectric layer FET (DM-below-NCFET) for biomolecules

3 Simulation methodology

Visual TCAD system is used for simulating all results. This region is confined to Bio1 and Bio2 to introduce and identify biomolecules. In this work, impartial(neutral) biomolecules, for example, streptavidin, protein, uricase, and ChO_x are created in the gap of the cavity. Drain voltage, V_d (V) is fixed in the entire process at 0.4 V. Gate voltage, V_{gs} is different from 0 to 4 V. temperature (T) is fixed during the entire device simulation at 300 K (room temperature). So, the existence of impartial biomolecules can be patterned by initiating substance with a dielectric constant related to impartial biomolecules. This method has been calibrated with the outcomes of the experiments [6, 21].

The design is used in concentration-dependent mobility (CONMOB), field-dependent drift velocity (FLDMOB) model, Shockley–Read–Hall (SRH) recombination model, and polarization model. We utilize visual TCAD multiphysics to play out every single mathematical reproduction. Bioelectronics is focused on straightforwardly coupling biomolecular work units of high-atomic weight and incredibly confounded subatomic construction (chemicals, receptors, or entire, cells, and so on just as tissues, cell clusters, and organisms) with electronic or optical transducer device. Elective and new ideas are constantly emerged and proposed for future data advancements to address, control, read, and use data, requiring the improvement of constructions for signal take-up, transduction, enhancement, handling, and change [24, 25].

The absorption of various biomolecules is patterned by the initiation of an non-conductor having the identical dielectric constant such as a specific biomolecule into the cavity's gap. The void cavity (i.e., without biomolecules present) is demonstrated by applying an insulator with dielectric constant (K = 1), although the dielectric constant of the diverse studied biomolecules is like this as shown in Table 1: protein (K = 8) [26], uricase (K = 1.5), ChO_x (K = 3.5), and streptavidin (K = 2.1) [27]. Streptavidin is utilized to distinguish Marek's disease virus (MDV) applying an enzyme-linked immunosorbent assay (ELISA) method [28, 29].

The essential thought behind a MD simulation is clear. Given the places of the multitude of atom in a biomolecular framework (e.g., a protein encompassed by water and maybe a liquid bilayer), one can compute the force applied on every particle by the wide range of various molecules. One can hence utilize Newton's laws of movement to foresee the spatial place of every molecule as an element of time. Specifically, one stages through time, more than once ascertaining the force on every iota and afterward utilizing those force to refresh the position and speed of every particle. The subsequent direction is, basically, a three-layered film that depicts the nuclear level setup of the framework at each point during the simulated time interval [30].

4 Results and discussion

4.1 Impact of dielectric modulation for cavity gap 10 nm

Figure 2 shows the curve between drain current, I_d (A) in linear scale and log scale vs gate voltage, V_{gs} (V) of different biomolecules for double metal below negative capacitance field-effect transistor (DM-below-NCFET) at V_d = 0.4 V, channel length, L_c = 20 nm, and oxide thickness T_o = 1 nm. The drain current of protein (K = 8) is higher than the others biomolecules, as enhanced drain current with an increase in dielectric constant that reflects the improved gate coupling capacitance and lowered leakage current I_{off} in the left side of the graph.

Table 1 The value of dielectric constant for biomolecule

Biomolecules	Dielectric constant
Uricase	1.5
Streptavidin	2.1
ChO_x	3.5
Protein	8

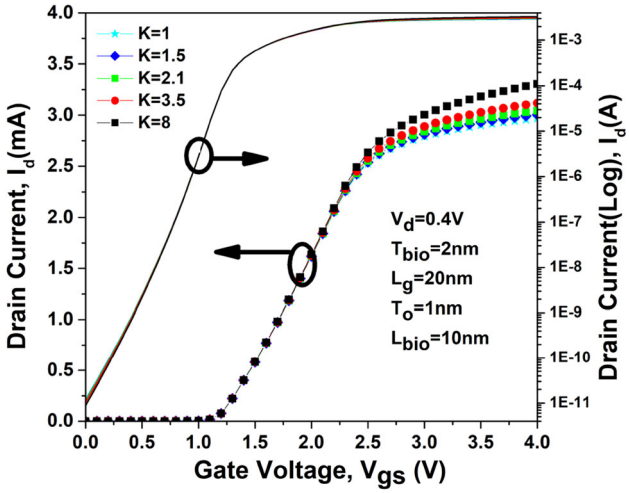


Fig. 2 Transfer characteristic of DM-below-NCFET for different biomolecules at drain voltage 0.4 V

Figure 3 reveals that transconductance is a factor of the shift in drain current (I_d) to shift in gate voltage (V_{gs}) at drain voltage (0.4 V). The transconductance curve of protein is higher than the streptavidin, uricase, ChO_x, without molecules. Thus, the g_m has taken from the curve drain current vs gate voltage that reflects enhanced gate control and lowered short channel effects (SCEs) and also increases average

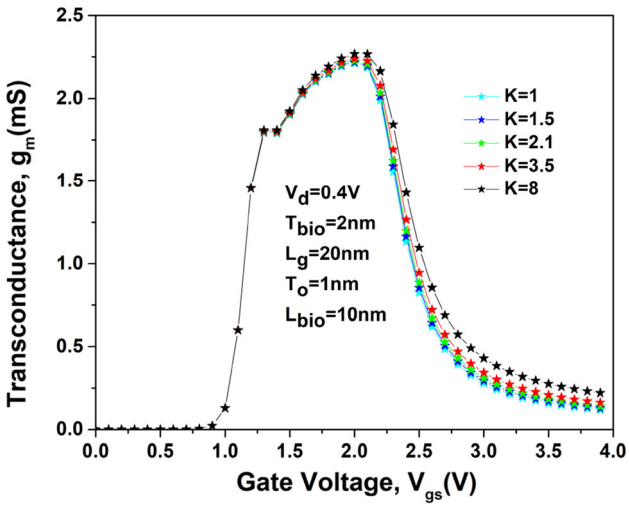


Fig. 3 Transconductance (g_m) of DM-below-NCFET for different biomolecules at V_d 0.4 V

carrier velocity, upgraded electron mobility with increase transconductance [31].

Figure 4 shows the threshold voltage, V_{th} (V) of protein is enhanced than other biomolecules configuration at drain voltage, V_d = 0.4 V because of enlarged gate control over the channel and biomolecule of protein is the better shield of drain-side potential so that improved the character of the threshold voltage. The threshold voltage is directly related to the dielectric constant was applied as a sensing application for recognition. Figure 5a shows the switching ratio, I_{on}/I_{off} of protein is higher than the different biomolecules at L_g = 20 nm, T_o = 1 nm, L_{bio} = 10 nm, and V_d = 0.4 V, decreased leakage current, I_{off} (A/um), and high switching speed for protein molecules (K = 8) is also obtained. Figure 5b implies the sensitivity of various biomolecules for instance protein, streptavidin, uricase, ChO_x, and without biomolecules at drain voltage 0.4 V as shown in Table 2. The cavity gap is completely filled with dielectric constant. The formula for sensitivity is given in Eq. (1).

$$S = \frac{I_{on}(K > 1)}{I_{on}(K = 1)} \quad (1)$$

Figure 6a–e represents the contour plot of electronic concentration of channel from source to drain for various dielectric constant Fig. 6a implies biomolecule is created, the extra charges become caused in the channel from source to drain in the region beneath the cavity. Charge quantity caused hangs on the various biomolecules. The electronic concentration of protein at drain is worked at lower concentration. Due to the smallest value of the dielectric constant of without molecules, extra charges are created in the channel as shown in Fig. 6f that is deducted from the increase of drain current.

Figure 7a–e reflects the contour plot of the potential of the channel for diverse biomolecules. It is also

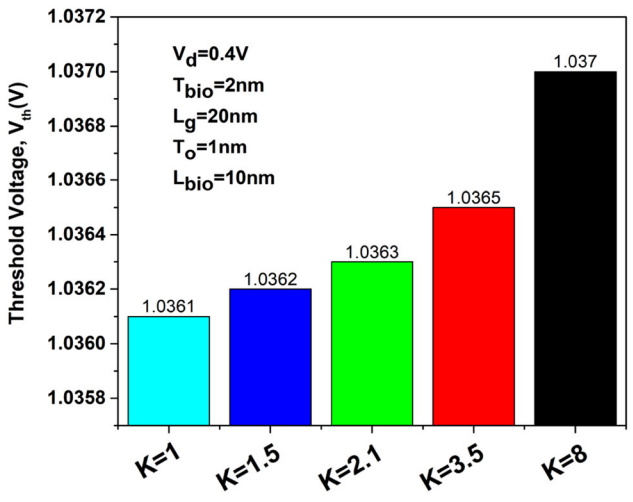


Fig. 4 The threshold voltage (V_{th}) of DM-below-NCFET for different biomolecules at V_d 0.4 V

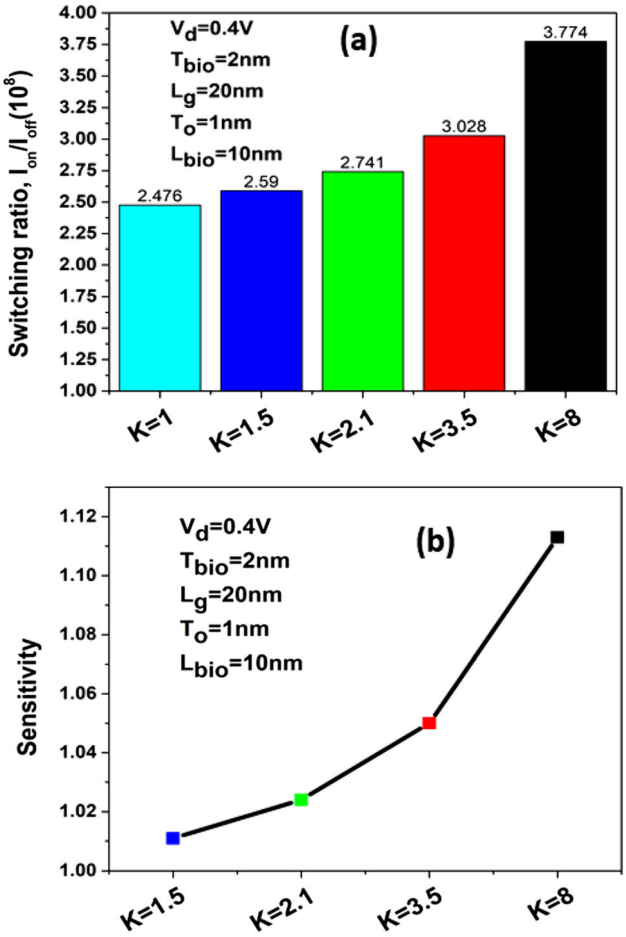


Fig. 5 **a** Switching ratio (I_{on}/I_{off}), **b** sensitivity of DM-below-NCFET for different biomolecules at drain voltage 0.4 V

Table 2 Value and parameter of DM-Below-NCFET for different biomolecules at T_o = 1 nm and cavity length 10 nm

Parameter	$K = 8$	$K = 3.5$	$K = 2.1$	$K = 1.5$	$K = 1$
I_{on} (mA)	3.31	3.12	3.04	3.01	2.97
I_{off} (pA)	8.76	10.3	11.1	11.6	12
I_{on}/I_{off} (10^8)	3.77	3.03	2.74	2.59	2.48
V_{th} (V)	1.037	1.0365	1.0363	1.0362	1.0361
g_m (mS)	2.27	2.23	2.21	2.2	2.19
g_d (nS)	0.662	0.698	0.718	0.731	0.744

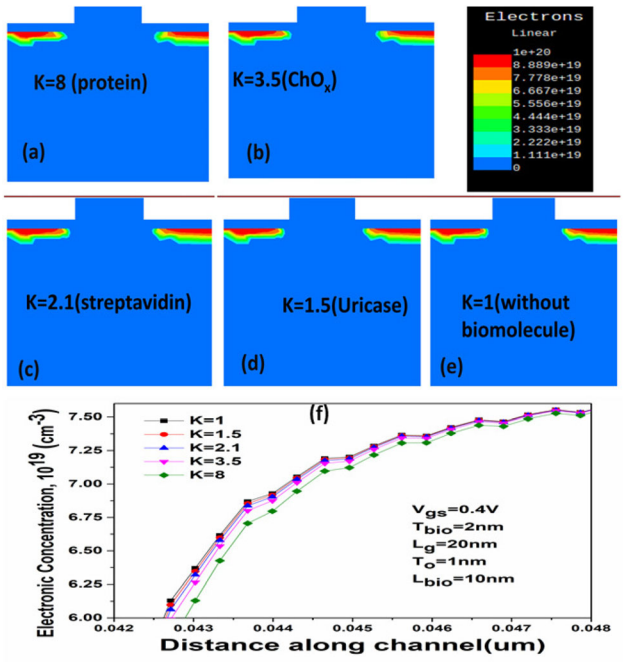


Fig. 6 **a**–**e** Electronic concentration contour plot of DM-below-NCFET at V_d 0.4 V for different biomolecules, for example, **a** K = 8, **b** K = 3.5, **c** K = 2.1, *d* K = 1.5, *e* K = 1, and **f** graph of electronic concentration vs distance along channel

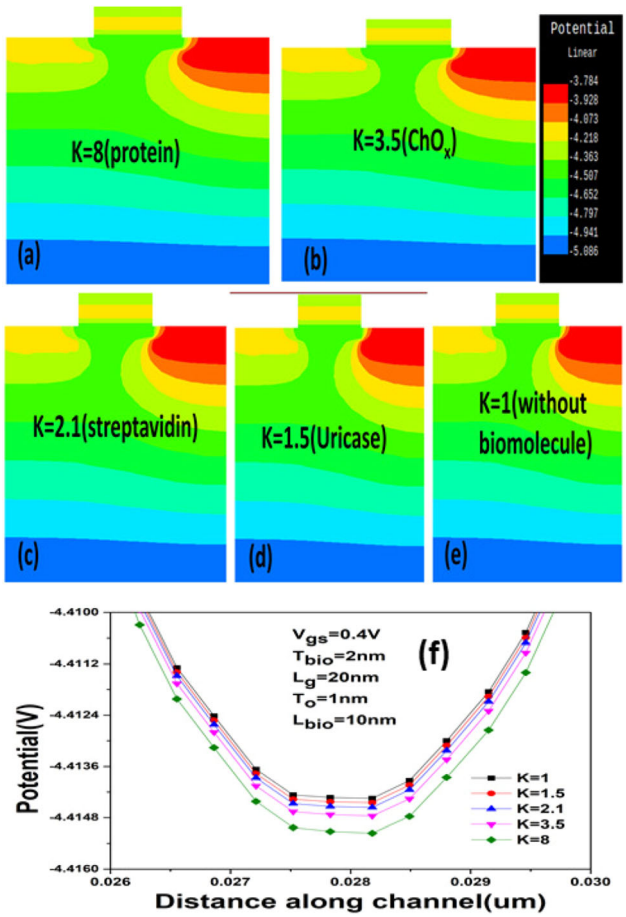


Fig. 7 Potential contour plot of DM-below-NCFET at V_d 0.4 V for different biomolecules for example **a** K = 8, **b** K = 3.5, **c** K = 2.1, **d** K = 1.5, **e** K = 1, **f** Potential graph of DM-below-NCFET at drain voltage 0.4V for different biomolecules

the electrical performance of DM-below-NCFET. Changes of the potential occur (at the drain side) under the cavity region. When the cavity gap is immobilized by biomolecules [32], a deformation in potential is detected because of the shift in the dielectric constant caused by the biomolecules as shown in Fig. 7f. This change in potential can as well be applied to identify the existence of biomolecules in the cavity gap area.

Figure 8 signifies the plot of drain current, I_d (A) vs drain voltage, V_d (V) at gate voltage V_{gs} (0.4) oxide thickness of 1 nm, cavity gap of 10 nm, and channel length of 20 nm. Output conductance is the driving capacity of the device as characterized mirrors the locale of device activity as shown in Fig. 9. At first, a high-output conductance in the linear region is observed with increasing drain voltage beyond pinch-off voltage owing to drain-induced barrier lowering (DIBL) as well as channel length modulation (CLM) [33]. The g_d

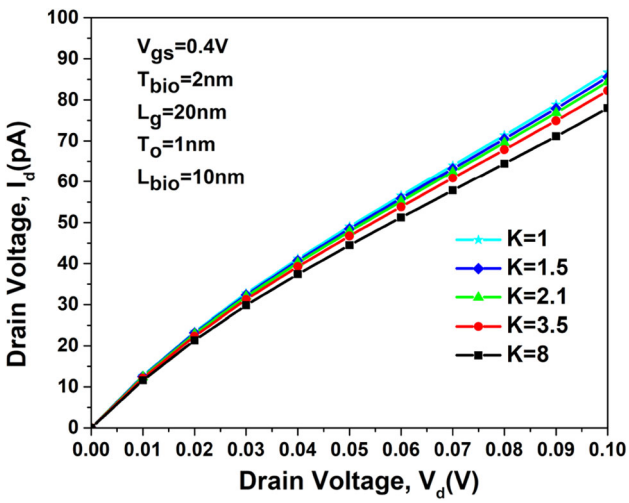


Fig. 8 Graph of drain current vs drain voltage of DM-below-NCFET at gate voltage 0.4 V for different biomolecules

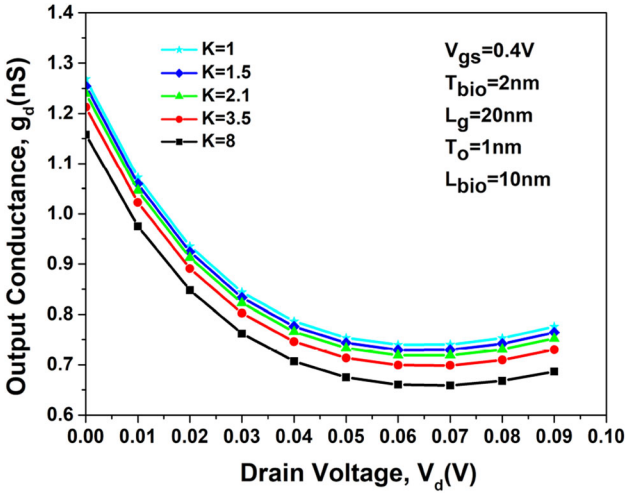


Fig. 9 Graph of output conductance vs drain voltage of DM-below- NCFET at gate voltage 0.4 V for different biomolecules

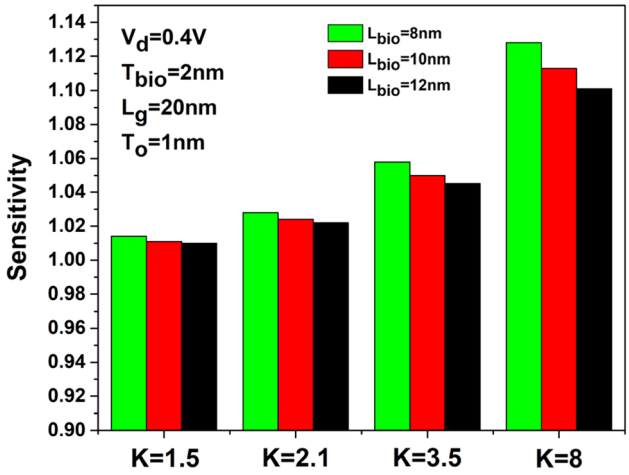


Fig. 11 Sensitivity of DM-below-NCFET at drain voltage 0.4 V for various biomolecules with different cavity length L_{bio} = 8 nm, 10 nm, and 12 nm

4.2 Effect of length modulation in cavity gap

Figure 10 shows graph of drain current vs gate voltage for various cavity length L_{bio} = 8 nm, 10 nm, and 12 nm with various biomolecules for instance ChO_x, protein, streptavidin, uricase, and without biomolecules. The drain current of protein with cavity length (L_{bio} = 8 nm) is enhanced over others as the dielectric constant increases with lower cavity reflects to improve drain current.

Figure 11 implies that the switching ratio was obviously lowered due to a rise in the leakage current subsequent from greater capacitance with the rise in the cavity gap length. Furthermore, the sensitivity of the DM-below-NCFET biosensor with distinct cavity gap lengths was too calculated, exposing that it improved meaningfully with the rise in L_{bio} (Fig. 12). Since this figure, it is stated that the sensitivity enhanced with a dielectric constant for offered L_{bio}. Thus, for L_{bio} = 12 nm, the nanogap-inserted DM-below-NCFET biosensor indicated improved sensitivity for protein (K = 8) in analogy with the additional biomolecules.

Figure 13 gives the plot of limit of detection (LOD) vs different K values with standard deviation 0.000719 for K = 1, 0.000719 for K = 1.5, 0.00072 for K =

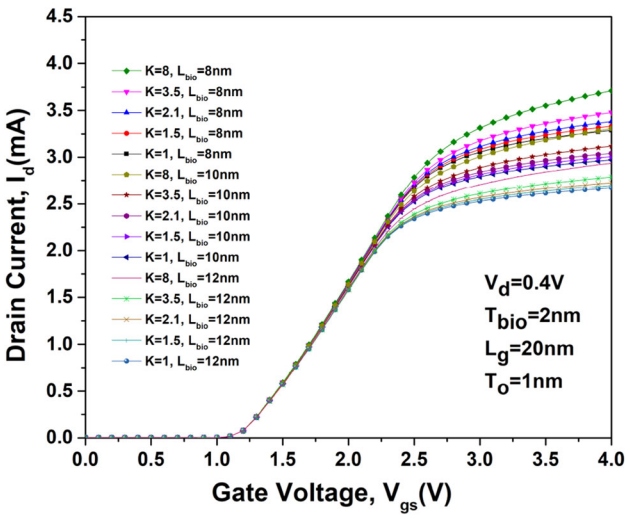


Fig. 10 Plot of drain current vs gate voltage of DM-below-NCFET at drain voltage 0.4 V for various biomolecules with different cavity length L_{bio} = 8 nm, 10 nm, and 12 nm

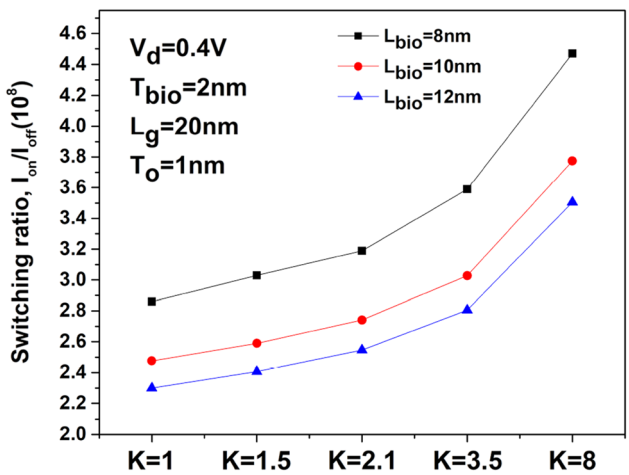


Fig. 12 Switching ratio of DM-below-NCFET at drain voltage 0.4 V for various biomolecules with various cavity length L_{bio} = 8 nm, 10 nm, and 12 nm

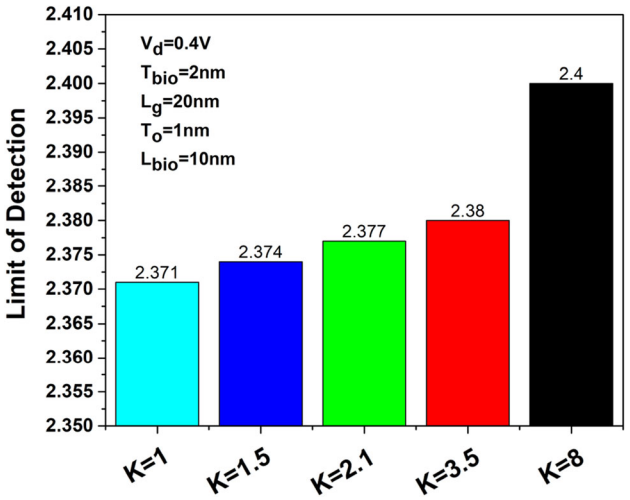


Fig. 13 Plot of limit of detection (LOD) vs different K values with standard deviation 0.000719 for $K = 1$, 0.000719 for $K = 1.5$, 0.00072 for $K = 2.1$, 0.000722 for $K = 3.5$, and 0.000727 for $K = 8$

2.1, 0.000722 for $K = 3.5$, and 0.000727 for $K = 8$. LOD of protein is higher than without molecule that indicating higher concentration of analyte in the sample that can be detected but not necessarily quantified [34]. It is determined by the response of standard deviation and slope. At the lower limit of detection in COVID-19 reflects it can miss more infected cases. These outcomes give understanding into the significance of controlling properties for amplifying sensitivity and limiting execution variety across device when planning and manufacturing nano-FET biosensors [35, 36].

Figure 14 shows the experimental data and simulated result with some parameters and reveals the

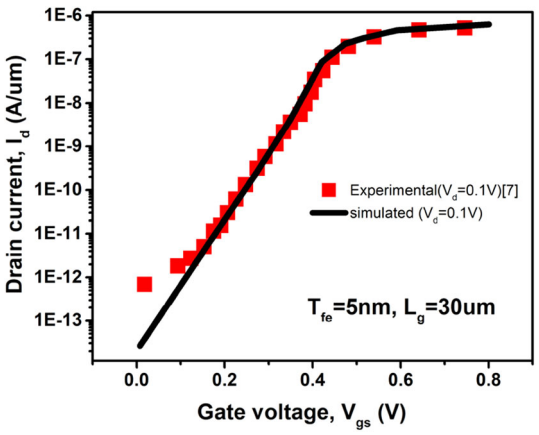


Fig. 14 Calibration results of experimental data and simulated data curve of drain current vs gate voltage at $V_d = 0.1$ and the channel length is 30 um

curve of drain current and gate voltage at drain voltage is 0.1 V, ferroelectric thickness is 5 nm, and the channel length is 30 um for negative capacitance field-effect transistor (NCFET), and data of experiment are taken from reference [7].

Organic FETs have formed into an astonishing area of examination and innovation to supplant exemplary inorganic semiconductors. Light-emitting diodes, organic photovoltaics, and thin-film transistors are as of now very much evolved and are at present being marketed an assortment of application for DM-below-NCFET. All the more as of late, organic FET has tracked down new applications in the field of biosensors [37, 38]. The FET-based biosensor turns into a promising contender for applications requiring ultra-sensitivity and fast reaction time. Moreover, new era complementary metal-oxide semiconductor (CMOS)-producing methods give advantages of scaling down, parallel detecting (for example, detecting clusters), and abilities to be coordinated with electronic circuits and frameworks. This would be a significant benefit for strong state-based biosensors to compete with other bio-detecting mechanism later on [23].

5 Conclusion

In this work, we examined the biosensing application of a double metal below negative capacitance FET (DM-below-NCFET). For the recognition of different biomolecules, the electrical performance was calculated in conditions of the shift of threshold voltage for protein enhanced by 0.09% over without biomolecule, the sensitivity of protein improved by 1.11 by times over $K = 1$, switching ratio of protein higher by 52% over to without biomolecule, the surface potential at cavity length of 10nm, as shown in Table 2, and measured higher limit of detection for protein by around 1.18% over without molecules. The DM-below-NCFET device was significantly improved, indicating better sensitivity for biomolecules with a large value of K. The proposed device realized that the sensitivity was greater for the recognition of protein ($K = 8$) in analogy with the various biomolecules. Moreover, the impacts of the gap of cavity length on the V_{vt}, sensitivity, and switching ratio(I_{on}/I_{off}) of the DM-below-NCFET biosensor were too analyzed, exposing that the sensitivity enhanced with the rise in the gap size although the whole biosensor

operation was marginally reduced. So, the gap of cavity length is well improved for improved execution, which will authorize the utilization of such cavity length of 8 nm high sensitivity, high-speed biosensors for recognition of several associated diseases.

Acknowledgements

The authors would like to thank Microelectronics Lab and Nano Bioelectronics Lab, Delhi Technological University, for giving essential facilities.

Author contributions

All authors added to the study's understanding and design.

Funding

Yash is grateful to the Council of Scientific Industrial Research (C.S.I.R.) for an award of CSIR-JRF (08/133(0050)/2020-EMR-I). BDM thanks the Science & Engineering Research Board (DST-SERB), Govt. of India for the award of Distinguished Fellowship (SB/DF/011/2019).

Data availability

The creators expressed above have every one of the suitable data associated with this examination work and will be dedicated to uncovering that they will be addressed to do as such future.

Declarations

Conflict of interest The writers broadcast that they have no known dispute of individual connections or interests that might have arisen to impact the work depicted in this article.

Ethical standards The authors have found all the moral standards and will intended to follow them in future.

Consent to participate and publication Since the related research document is created on the 'no-life science journal.' Therefore, 'Non-Applicable' at this point. Still, the authors have turned over all journal guidelines and consent to the agencies for additional processing.

References

1. H.M. Ahmed, R. Chaujar et al., Rapid detection of biomolecules in a dielectric modulated gan moshemt. J. Mater. Sci. **31**(19), 16609–16615 (2020)
2. A. Kumar, M. Roy, N. Gupta, M. Tripathi, R. Chaujar, Dielectric modulated transparent gate thin film transistor for biosensing applications. Mater. Today **28**, 141–145 (2020)
3. W. Withayachumnankul, K. Jaruwongrungsee, A. Tuantranont, C. Fumeaux, D. Abbott, Metamaterial-based microfluidic sensor for dielectric characterization. Sens. Actuators A **189**, 233–237 (2013)
4. J.S. Lee, K.-Y. Shin, O.J. Cheong, J.H. Kim, J. Jang, Highly sensitive and multifunctional tactile sensor using free-standing zno/pvdf thin film with graphene electrodes for pressure and temperature monitoring. Sci. Rep. **5**(1), 1–8 (2015)
5. P. Bergveld, The development and application of fet-based biosensors. Biosensors **2**(1), 15–33 (1986)
6. H. Im, X.-J. Huang, B. Gu, Y.-K. Choi, A dielectric-modulated field-effect transistor for biosensing. Nat. Nanotechnol. **2**(7), 430–434 (2007)
7. M. Lee, P.-G. Chen, C. Liu, K. Chu, C.-C. Cheng, M.-J. Xie, S.-N. Liu, J.-W. Lee, S.-J. Huang, M.-H. Liao, et al., Prospects for ferroelectric hfzrox fets with experimentally cet = 0.98 nm, ssfor = 42mv/dec, ssrev= 28mv/dec, switch-off < 0.2 v, and hysteresis-free strategies. in *2015 IEEE International Electron Devices Meeting (IEDM)*. IEEE (2015), pp. 22–25
8. X.T. Vu, J.F. Eschermann, R. Stockmann, R. GhoshMoulick, A. Offenhäusser, S. Ingebrandt, Top-down processed silicon nanowire transistor arrays for biosensing. Physica Status Solidi **206**(3), 426–434 (2009)
9. A. Kumar, M. Tripathi, R. Chaujar, Ultralow-power dielectric-modulated nanogap-embedded sub-20-nm TGRC-MOSFET for biosensing applications. J. Comput. Electron. **17**(4), 1807–1815 (2018)
10. K. Moselund, D. Bouvet, V. Pott, C. Meinen, M. Kayal, A. Ionescu, Punch-through impact ionization MOSFET (PIMOS): from device principle to applications. Solid-State Electron. **52**(9), 1336–1344 (2008)
11. C.-I. Lin, A.I. Khan, S. Salahuddin, C. Hu, Effects of the variation of ferroelectric properties on negative capacitance FET characteristics. IEEE Trans. Electron Devices **63**(5), 2197–2199 (2016)

12. G. Pahwa, T. Dutta, A. Agarwal, Y.S. Chauhan, Physical insights on negative capacitance transistors in nonhysteresis and hysteresis regimes: MFMIS versus MFIS structures. IEEE Trans. Electron Devices **65**(3), 867–873 (2018)
13. G. Pahwa, A. Agarwal, Y.S. Chauhan, Numerical investigation of short-channel effects in negative capacitance MFIS and MFMIS transistors: subthreshold behavior. IEEE Trans. Electron Devices **65**(11), 5130–5136 (2018)
14. B. Awadhiya, P.N. Kondekar, S. Yadav, P. Upadhyay, Insight into threshold voltage and drain induced barrier lowering in negative capacitance field effect transistor. Trans. Electr. Electron. Mater. **22**(3), 267–273 (2021)
15. Y. Pathak, B.D. Malhotra, R. Chaujar, TCAD analysis and simulation of double metal negative capacitance FET (DM NCFET). in *2021 Devices for Integrated Circuit (DevIC)*. IEEE (2021), pp. 224–228. IEEE
16. M. Sharma, R. Chaujar, Design and investigation of recessed-t-gate double channel HEMT with INGAN back barrier for enhanced performance. Arab. J. Sci. Eng. 1–8 (2021)
17. B. Kumar, R. Chaujar, Analog and RF performance evaluation of junctionless accumulation mode (JAM) gate stack gate all around (GS-GAA) finfet. Silicon **13**(3), 919–927 (2021)
18. N. Gupta, R. Chaujar, Optimization of high-k and gate metal workfunction for improved analog and intermodulation performance of gate stack (GS)-GEWE-SINW MOSFET. Superlatt. Microstruct. **97**, 630–641 (2016)
19. J. Madan, R. Gupta, R. Chaujar, Mathematical modeling insight of hetero gate dielectric-dual material gate-GAA-tunnel FET for VLSI/analog applications. Microsyst. Technol. **23**(9), 4091–4098 (2017)
20. J. Madan, R. Gupta, R. Chaujar, Performance investigation of heterogeneous gate dielectric-gate metal engineered-gate all around-tunnel FET for RF applications. Microsyst. Technol. **23**(9), 4081–4090 (2017)
21. R. Mann, R. Chaujar, TCAD investigation of ferroelectric based substrate MOSFET for digital application. Silicon 1–10 (2021)
22. Y. Pathak, B.D. Malhotra, R. Chaujar, A numerical study of analog parameter of negative capacitance field effect transistor with spacer. in *2021 7th International Conference on Signal Processing and Communication (ICSC)*. IEEE (2021), pp. 277–281
23. Y.-C. Syu, W.-E. Hsu, C.-T. Lin, Field-effect transistor biosensing: devices and clinical applications. ECS J. Solid State Sci. Technol. **7**(7), 3196 (2018)
24. C.G. Siontorou, F.A. Batzias, V. Tsakiri, A knowledge-based approach to online fault diagnosis of FET biosensors. IEEE Trans. Instrum. Meas. **59**(9), 2345–2364 (2009)
25. S. Kumar, M. Umar, A. Saifi, S. Kumar, S. Augustine, S. Srivastava, B.D. Malhotra, Electrochemical paper based cancer biosensor using iron oxide nanoparticles decorated PEDOT: PSS. Anal. Chim. Acta **1056**, 135–145 (2019)
26. S. Kim, J.S. Ho, L.Y. Chen, A.S. Poon, Wireless power transfer to a cardiac implant. Appl. Phys. Lett. **101**(7), 073701 (2012)
27. S. Busse, V. Scheumann, B. Menges, S. Mittler, Sensitivity studies for specific binding reactions using the biotin/streptavidin system by evanescent optical methods. Biosens. Bioelectron. **17**(8), 704–710 (2002)
28. I. Davidson, M. Malkinson, C. Strenger, Y. Becker, An improved ELISA method, using a streptavidin-biotin complex, for detecting Marek's disease virus antigens in feathertips of infected chickens. J. Virol. Methods **14**(3–4), 237–241 (1986)
29. H. Mulaosmanovic, E.T. Breyer, S. Dünkel, S. Beyer, T. Mikolajick, S. Slesazeck, Ferroelectric field-effect transistors based on hfo2: a review. Nanotechnology (2021)
30. S.A. Hollingsworth, R.O. Dror, Molecular dynamics simulation for all. Neuron **99**(6), 1129–1143 (2018)
31. V. Narendar, K.A. Girdhardas, Surface potential modeling of graded-channel gate-stack (GCGS) high-k dielectric dual-material double-gate (DMDG) MOSFET and analog/RF performance study. Silicon **10**(6), 2865–2875 (2018)
32. S. Nara, R. Kandpal, V. Jaiswal, S. Augustine, S. Wahie, J.G. Sharma, R. Takeuchi, S. Takenaka, B.D. Malhotra, Exploring providencia rettgeri for application to eco-friendly paper based microbial fuel cell. Biosens. Bioelectron. **165**, 112323 (2020)
33. S.I. Amin, R. Sarin, Charge-plasma based dual-material and gate-stacked architecture of junctionless transistor for enhanced analog performance. Superlatt. Microstruct. **88**, 582–590 (2015)
34. D. Martens, P. Bienstman, Study on the limit of detection in MZI-based biosensor systems. Sci. Rep. **9**(1), 1–8 (2019)
35. J. Li, Y. Zhang, S. To, L. You, Y. Sun, Effect of nanowire number, diameter, and doping density on nano-FET biosensor sensitivity. ACS Nano **5**(8), 6661–6668 (2011)
36. Y. Cui, Q. Wei, H. Park, C.M. Lieber, Nanowire nanosensors for highly sensitive and selective detection of biological and chemical species. Science **293**(5533), 1289–1292 (2001)
37. L. Kergoat, B. Piro, M. Berggren, G. Horowitz, M.-C. Pham, Advances in organic transistor-based biosensors: from organic electrochemical transistors to electrolyte-gated organic field-effect transistors. Anal. Bioanal. Chem. **402**(5), 1813–1826 (2012)
38. I. Novodchuk, M. Bajcsy, M. Yavuz, Graphene-based field effect transistor biosensors for breast cancer detection: a review on biosensing strategies. Carbon **172**, 431–453 (2021)

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

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Detection of Defective Steel Surface with Image Segmentation

Chirag Gupta
Department of Mechanical Engineering
Delhi Technological Universty
Delhi, India
chiraggupta2506@gmail.com

Gaurav Chauhan
Department of Mechanical Engineering
Delhi Technological Universty
Delhi, India
chauhangaurav9721@gmail.com

Hari Dev S Aiyar
Department of Mechanical Engineering
Delhi Technological University
Delhi, India
aiyarhari@gmail.com

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760806

***Abstract*— The detection of defects is an important part of steel manufacturing. Conventional ways of detecting defects in steel include ultrasonic testing, magnetic powder testing, osmosis testing, eddy current testing and X-ray testing. In this study, we used deep learning and computer vision to identify surface defects on steel. The study used U-Net, Mask R-CNN and Xception-V1 which are deep neural network models used to detect specific parts of an image or image segmentation. The Xception-V1 model performed well with complex defects whereas Mask-RCNN gave better results in rare defects. U-Net performed well with every type of defect and was a versatile model. The accuracy of U-Net model came out to be 80.21%. Images with no defects resulted in higher percentage of false predictions. It can be concluded that the existing models, while are capable of identifying different defects, they are not very good at predicting if there is no defect in this particular case.**

Keywords*— *Defect Detection, Deep Learning, Image Segmentation

I. INTRODUCTION

Steel is a vital component of our daily life. Steel is present in our daily lives in some shape or another. It may be found in everyday tools, utensils, pipes, automobiles, military weaponry, machineries, and electrical towers, to name a few examples. It is a carbon-iron alloy which comprises of 1% manganese, less than 2% carbon and trace quantities of oxygen, sulphur, phosphorus and silicon. It is the most significant and versatile substance available. Steel's strength and inherent usefulness served as the backbone of industrialised economies [1]. In comparison to other materials of its type, it has cheaper production costs. The characteristics of steel which makes it so useful and adaptable are:

- Corrosion resistance
- Heat resistance
- Hard and wear resistance
- Machinability
- Weldability
- Hot and cold formability

Steel, like all other materials, is susceptible to flaws. Defects can arise as a result of design and manufacturing equipment failures, as well as bad working conditions. Surface defects on steel during manufacturing include [2][3][4][5][6][7]:

1) Cracking: It is the most prevailing type of defect in steel. They occur in various shapes and sizes. A crack maybe longitudinal, transversal, crater cracks, branching crack or stress-corrosion crack.

2) Surface Porosity: It is formed when gas breaks the steel surface when trying to escape.

3) Pitting: It is a form of extremely localized corrosion which results in formation of small "pits" on the surface of metal.

4) Dents or depressions: These are formed on the surface due to hard blows or excessive pressure applied.

As a result, defect detection is a crucial aspect of steel manufacturing that manufacturers must have in order to increase the quality of the created goods while minimising the impact on production. Defect-detection technology, on the product surface, relates to the identification of spots, pits, scratches, colour variances, and defects. Current defect-detection techniques in steel includes processes such as magnetic particle testing [8], liquid penetrant test, eddy current testing [9], osmosis testing [10] and ultrasonic testing [11].

Magnetic particles can be detected by following methods: dry magnetic particle testing (DMPT) and wet magnetic particle testing (WMPT). Magnetic powder is combined with oil, water, or another liquid medium for wet magnetic particle detection. Due to hydraulic pressure and the attraction caused by the external magnetic flux, magnetic powder identifies the position of flaws. The liquid medium is recyclable, and the approach has a high sensitivity. The magnetic powder is directly bonded to the surface of the magnetised workpiece for fault identification in dry magnetic particle testing. It's utilised to evaluate flaws in segments that aren't suited for wet detection on a local level. Magnetic powder testing precision is influenced by the profile and raggedness of the test piece, the magnetization method employed, the geometrical features of faults, and operator's expertise.

LPI (liquid penetrant inspection) is utilized in order to locate surface-breaking faults in non-porous materials (plastics, metals, ceramics). It is a less expensive test method used to detect additional defects on surface of metal such as leakage of new products, hairline cracks and fatigue cracks in in-service components in welding, casting, and forging. In order to clean and dry surface-breaking faults through capillary action in dry penetrant inspection (DPI), a low-surface-tension fluid is supplied. Once sufficient login time has been allowed and over-entry is removed by an error-pulling engineer, which allows indicators to be displayed.

Electromagnetic induction is utilised in eddy current testing to detect surface and subsurface defects. An alternating electrical current excites a coiled copper wire, creating a magnetic field around it. When the coil comes close to a conductive substance, eddy currents — currents that are the polar opposite of those in the coil — are generated in the material. Defects in the metal area interfere with the flow of eddy currents, leading to a magnetic field that alters the balance of the system, which can be measured by monitoring coil fluctuations. The profile and material of the test piece, as

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:47 UTC from IEEE Xplore. Restrictions apply.

well as the kind and characteristics of the coil, all impact the accuracy of eddy current detection.

In ultrasonic testing, high frequency sound waves are used to locate cracks on metal surface. The ultrasound waveform may be received in form of reflection or attenuation. Ultrasonic testing is influenced by the angle formed between the acoustic propagation direction and the fault surface. To limit leakage detection, the proper detection sensitivity and associated probe must be used. Projection, direction, sound contact quality, instrument operating frequency, and probe efficacy are all aspects that influence ultrasonic testing.

Traditional defect-detection methods offer their own set of benefits. Traditional detection methods are quite specific. The osmosis testing method, for example, is only appropriate for finding faults in extremely permeable and non-porous materials. Traditional detection methods, on the other hand, require manual intervention to accomplish. Before testing, some time is spent on instrument debugging, and the cost of equipment development is significant. The equipment life and production precision of conventional procedures are limited as a result. Defect-detection technology that is automated offers clear advantages over manual detection. Computer vision is extremely precise and efficient. Automating the defect-detection process may significantly cut manufacturing costs and time, enhance production efficiency, and ultimately improve product quality.

Because of their versatility and lack of reliance on human aid, machine vision and deep-learning techniques have become one of the most essential technologies for automating defect diagnosis in recent years [12][13]. Although computer vision provides superior inspection outcomes and reduced costs, it relies on enormous volumes of learnt data to update models and increase inspection accuracy. Image capture, flaw detection, and classification are all part of computer vision detection. Machine vision is frequently employed due to its rapid, non-destructive, accurate, and low of cost qualities. The colour, texture, and geometric properties of things are used by machine vision to identify them. The difficulty of image processing is determined by quality and versatility of images collected. Consequently, the rate of error detection and accuracy of defect classification and detection depends on the quality of the image processing technique that has been used. Deep learning is a defect-detection approach based on image processing. To extract important qualities from enormous volumes of data, image processing is used on a wide scale.

Image segmentation is the process of in which the image is partitioned into multiple segments so that its objects and boundaries can be located in the image [14]. Like videos and audios, images are used to transfer data. They contain various useful information. Extracting information and thus understanding images to accomplish some works is important. It is the first step in understanding the image is the image segmentation. Image segmentation results in a set of contours or masks that are extracted from image. It may be divided into semantic segmentation and instance segmentation.

II. METHODS

A. U-NET

U-NET is a semantic segmentation model which can perform localization. It was initially built for bio medical image segmentation [15]. The U-NET uses encoder decoder architecture. The input image undergoes many convolution and deconvolution operations and the resulting output image is the mask (defect indicator) which happens to be of the same size as the input image. The encoder performs the convolution and max pooling operation which reduces the resolution of the image. While encoding, the image size reduces and the depth of image increases. This helps the network to understand "what" the information is but loses sight of "where" exactly the information is. The decoder upsamples the image using deconvolution layer and extracts the intricate details such as precise location of information in the dataset using convolutional layers. The model output is the image with 256x1600 resolution.

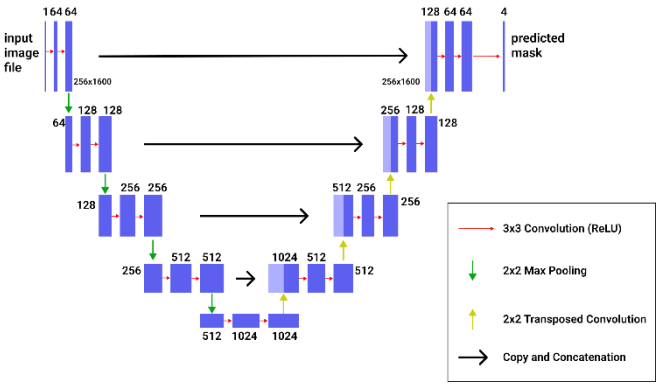


Fig. 1. U-Net Architecture

B. MASK-RCNN

Mask RCNN is an instance segmentation model which can perform localization and classification at the same time [16]. FPN (Feature Pyramid Network) extracts the features. RPN (Region Proposal Network) scans over the FPN features to suggest ROI (Region of Interest). The ROI Align classifies the object by using SVM classifier, in which each type of defect is trained separately, and is passed on to a mask generator which gives the region of defect. The backbone of the model has been thus chosen as FPN [17]. MaskRCNN contains a lot of parameter and comparatively takes huge amount of time to train.

C. XCEPTION-V1

Xception is the extreme version of the Inception model developed by Google [18]. It no longer uses the standard inception modules and instead relies on depth wise separable convolutions. There are 14 modules for Xception model all of which except the first and last modules consist linear residual connections. In total it has 36 convolutional layers which function as feature extractors. The output from the last layer in the Xception model is upsampled and passed through a 1x1 convolution filter with sigmoid activation which predicts the mask.

III. DATASET

The dataset was obtained from one of Kaggle's competition "Severstal: Steel Defect Detection" [19]. It consists of 12568 images in training and 5506 images in testing. The dataset had the type of defect category and mask

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:47 UTC from IEEE Xplore. Restrictions apply.

(defect) for the training images. There were 4 types of unique defect. Since, no testing image had any defect mask we discarded them. The training set consisted of 6666 unique images having at least one defect. The images had a resolution of 256x1600x3 pixel.

We were left with 6666 images in the training set after the preprocessing. Majority of the defects belonged to Class 3 (77%). To ensure that the model does not create a bias, augmentation was performed on the dataset [20]. Blur, horizontal flip, vertical flip and rotation were applied to the original dataset as part of data augmentation.

IV. Experimental Setup

A. Model Training

The dataset obtained after preprocessing was split in the ratio of 0.85 for training set and 0.15 for validation set. MaskRCNN, U-NET and Xception V1 model and an ANN classifier were trained on the dataset to perform defect detection. MaskRCNN model used FPN backbone for the experiments. U-NET model was chosen with 5 encoder-decoder architecture. The mask was sent to a classifier for defect type prediction. Training was performed on Kaggle notebook with NVidia K80 GPUs and the model was built using Tensorflow framework. Batch size of 16 was chosen for training all the models in the experiment. Adam with default hyperparameters was used as the optimizer and training was completed when there was no improvement for the last 30 epochs.

B. Cost Function

Binary cross entropy cost function is used for the detection problem. It is used in binary classification problems. The function penalizes the prediction upon how different it is from the ground truth [21].

$$Loss = -\frac{1}{N}\sum_{i=1}^{n} y_i \cdot \log\left(p\left(y_i\right)\right) + \left(1 - y_i\right) \cdot \log\left(1 - p\left(y_i\right)\right) \quad (1)$$

Wherein, y_i is the ground truth & $p(y_i)$ is the prediction

C. Evaluation Metrics

The performance of the models were evaluated by using dice coefficient metric. Dice coefficient is comparable to precision in that it not only measures how many positives you uncover, but it also penalises you for false positives. The predicted mask is compared against the ground truth in terms of overlap of pixels. The formula is given by [22]:

$$DICE = \frac{2|X \cap Y|}{|X| + |Y|} \quad (2)$$

Wherein, X = predicted mask & Y = ground truth
$\cap$ = the intersection

V. Results

The metrics and losses in Fig. 2, Fig. 5 and Fig. 9 represent the performance of Xception, MaskRCNN and U-Net models respectively. The training stops when no improvement in the performance is noted after a certain number of epochs. MaskRCNN takes maximum amount of time for training due to its enormous number of parameters. Xception takes least amount of time. Pretrained weights have helped the model to start with a higher score. Data augmentation had helped the models to improve their performance.

In the Fig. 2, the Xception model achieved a dice coefficient score of 0.6887 in 27 epochs. The Xception model performed very well in predicting the masks of defect type 4 as observed in Fig. 3. In Fig. 4, the model mistook the dark grey lines on the sheet for defect class 2. This led to an increase in false predictions. It fared poorly in rest of the defects which were rare as observed.

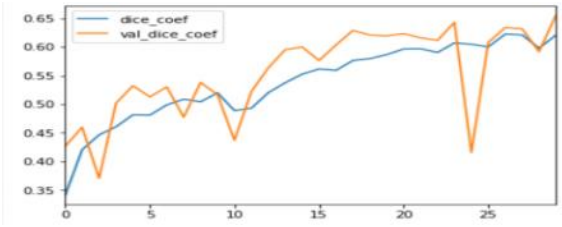


Fig. 2. Xception model dice score

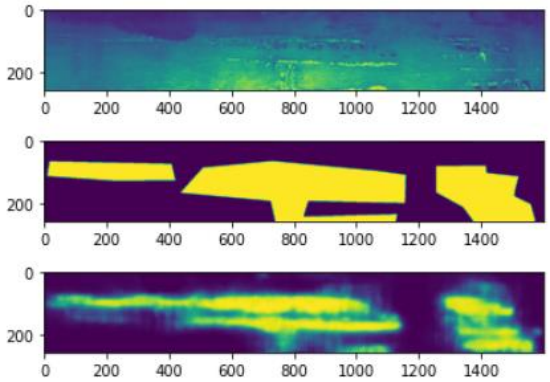


Fig. 3. Prediction of Xception model for class 4 defects

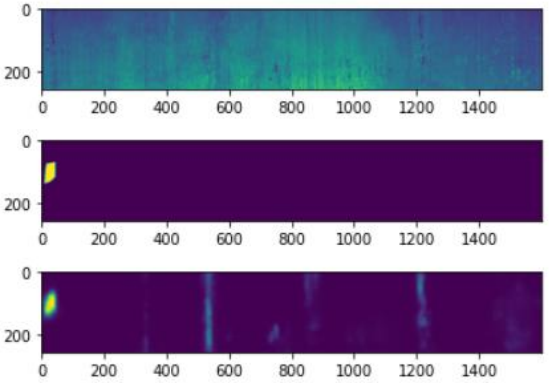


Fig. 4. Prediction of Xception model for class 1 defects and false prediction of class 2

The MaskRCNN model in the Fig. 5 has achieved a dice coefficient score of 0.7193 in 19 epochs. MaskRCNN suffers from the same problem as Xception did. Defect 2 false predictions were again visible in overwhelming numbers as described in Fig. 8. The model performed better than Xception in predicting the mask of defect type 1 and 2 (rare defects).

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:47 UTC from IEEE Xplore. Restrictions apply.

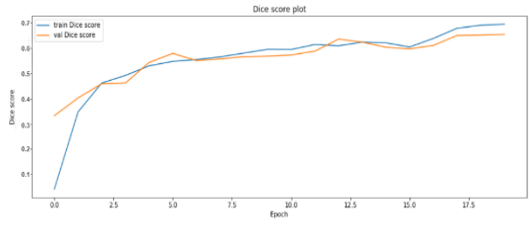


Fig. 5. MaskRCNN model dice score

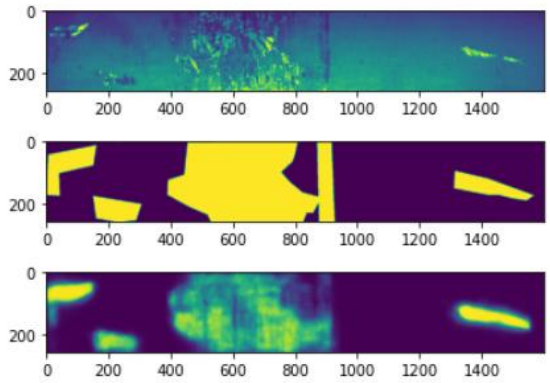


Fig. 6. Prediction of MaskRCNN model for class 4 and 3

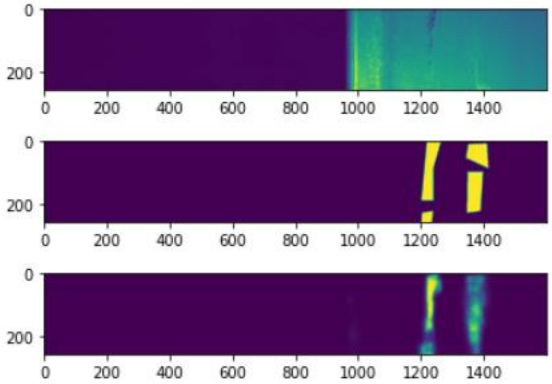


Fig. 7. Prediction of MaskRCNN model for class 1 and 2

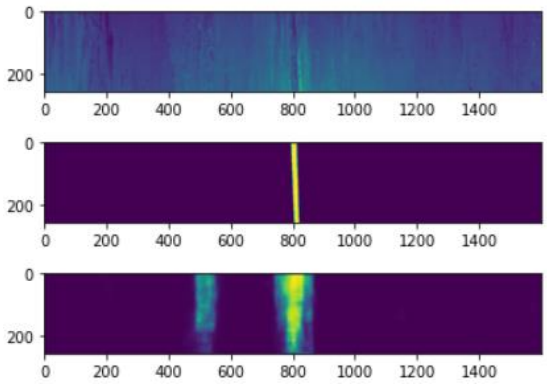


Fig. 8. False Prediction of MaskRCNN model for class 2

In Fig. 9, U-net model achieved a dice coefficient score of 0.8021 after 103 epochs. It stopped beforehand as there was no change in validation performance for the last 30 epochs. U-NET has the best accuracy amongst all the models. The predicted mask of defect type 1, 2 and 4 were the best among all the models (refer Fig. 10 & Fig. 11). But the model made defect prediction for no defect type class (Fig. 12). This model is feasible for real time deployment due to a smaller number of parameters.

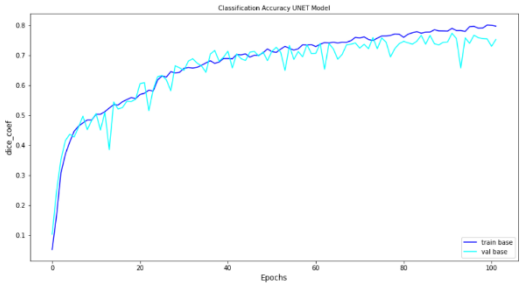


Fig. 9. U-NET model dice score

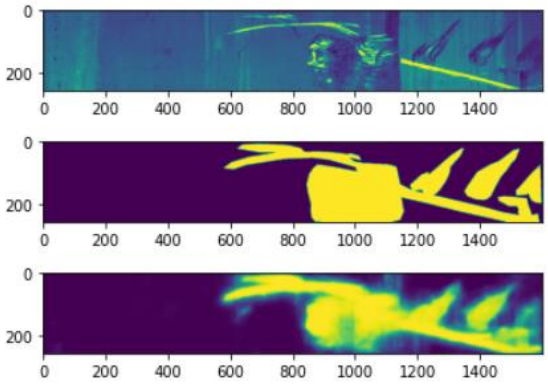


Fig. 10. Prediction of U-NET model for class 4

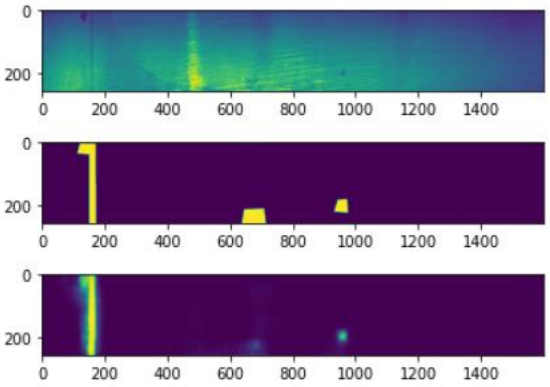


Fig. 11. Prediction of U-NET model for class 1 and 2

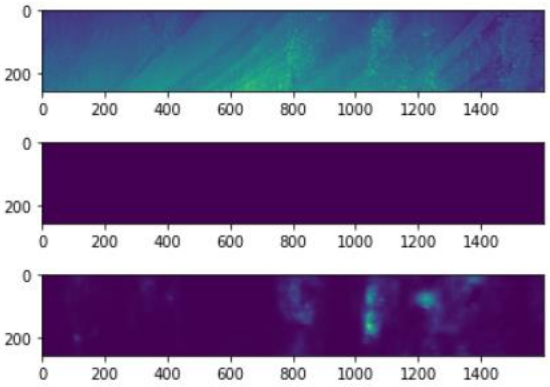


Fig. 12. Prediction of U-NET model for no class

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:47 UTC from IEEE Xplore. Restrictions apply.

VI. CONCLUSION

We have proposed a model to detect the defects found in the sheet metal using pre-trained models. We used Xception, MaskRCNN and U-NET to predict the masks and the defect class. Xception performed very well in complex defects, that is defects of type 4, which were substantial, but fared poorly in rare defects. MaskRCNN could handle the rare defects very well. U-NET was a versatile model as it could handle all the defects and most of the predicted masks were close to the ground truth. But it made false predictions for no defect type class. We also found that data augmentation and using pretrained weights helped in improving the metrics. As a part of future work, we would explore different segmentation algorithms and employ generative models to create synthetic data which will help the model to better cope with rare, complex and no defects category. Also, semi/weakly supervised learning approaches will be studied and implemented so that training data requirement can be reduced.

REFERENCES

[1] J. D. Glassman, A. Gomez, M. E. M. Garlock, and J. Ricles, "Mechanical properties of weathering steels at elevated temperatures," *J. Constr. Steel Res.*, vol. 168, 2020, doi: 10.1016/j.jcsr.2020.105996.

[2] Y. Peng, G. Liu, Y. Quan, and Q. Zeng, "The depth measurement of internal defect based on laser speckle shearing interference," *Opt. Laser Technol.*, vol. 92, 2017, doi: 10.1016/j.optlastec.2017.01.004.

[3] X. Xiao *et al.*, "Adipose stem cell-laden injectable thermosensitive hydrogel reconstructing depressed defects in rats: filler and scaffold," *J. Mater. Chem. B*, vol. 3, no. 27, 2015, doi: 10.1039/c5tb00270b.

[4] H. Kong, J. Yang, and Z. Chen, "Accurate and Efficient Inspection of Speckle and Scratch Defects on Surfaces of Planar Products," *IEEE Trans. Ind. Informatics*, vol. 13, no. 4, 2017, doi: 10.1109/TII.2017.2668438.

[5] J. shan Chen and C. sheng Li, "Prediction and Control of Thermal Scratch Defect on Surface of Strip in Tandem Cold Rolling," *J. Iron Steel Res. Int.*, vol. 22, no. 2, 2015, doi: 10.1016/S1006-706X(15)60017-6.

[6] I. G. Rodionova, A. I. Zaitsev, O. N. Baklanova, A. Y. Kazankov, V. V. Naumenko, and G. V. Semernin, "Effect of Carbon Steel Structural Inhomogeneity on Corrosion Resistance in Chlorine-Containing Media," *Metallurgist*, vol. 59, no. 9–10, 2016, doi: 10.1007/s11015-016-0173-2.

[7] J. Amaya, A. Lelah, and P. Zwolinski, "Design for intensified use in product–service systems using life-cycle analysis," *J. Eng. Des.*, vol. 25, 2014, doi: 10.1080/09544828.2014.974523.

[8] A. L. Elrefai and I. Sasada, "Magnetic Particle Detection System Using Fluxgate Gradiometer on a Permalloy Shielding Disk," *IEEE Magn. Lett.*, vol. 7, 2016, doi: 10.1109/LMAG.2016.2614248.

[9] G. D'Angelo, M. Laracca, S. Rampone, and G. Betta, "Fast Eddy Current Testing Defect Classification Using Lissajous Figures," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 4, 2018, doi: 10.1109/TIM.2018.2792848.

[10] A. Habib, C. H. Kim, and J. M. Kim, "A crack characterization method for reinforced concrete beams using an acoustic emission technique," *Appl. Sci.*, vol. 10, no. 21, 2020, doi: 10.3390/app10217918.

[11] M. Kusano *et al.*, "Mid-infrared pulsed laser ultrasonic testing for carbon fiber reinforced plastics," *Ultrasonics*, vol. 84, 2018, doi: 10.1016/j.ultras.2017.11.015.

[12] Y. Yang *et al.*, "A high-performance deep learning algorithm for the automated optical inspection of laser welding," *Appl. Sci.*, vol. 10, no. 3, 2020, doi: 10.3390/app10030933.

[13] Y. He, K. Song, Q. Meng, and Y. Yan, "An End-to-End Steel Surface Defect Detection Approach via Fusing Multiple Hierarchical Features," *IEEE Trans. Instrum. Meas.*, vol. 69, no. 4, 2020, doi: 10.1109/TIM.2019.2915404.

[14] Y. Song and H. Yan, "Image Segmentation Techniques Overview," 2018, doi: 10.1109/AMS.2017.24.

[15] W. Weng and X. Zhu, "INet: Convolutional Networks for Biomedical Image Segmentation," *IEEE Access*, vol. 9, 2021, doi: 10.1109/ACCESS.2021.3053408.

[16] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, 2020, doi: 10.1109/TPAMI.2018.2844175.

[17] I. Katsamenis, E. Protopapadakis, A. Doulamis, N. Doulamis, and A. Voulodimos, "Pixel-Level Corrosion Detection on Metal Constructions by Fusion of Deep Learning Semantic and Contour Segmentation," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2020, vol. 12509 LNCS, doi: 10.1007/978-3-030-64556-4_13.

[18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, 2017, vol. 2017-January, doi: 10.1109/CVPR.2017.195.

[19] Kaggle.com, "Severstal: Steel Defect Detection | Kaggle," 2020. https://www.kaggle.com/c/severstal-steel-defect-detection/data (accessed Oct. 28, 2021).

[20] C. Shorten and T. M. Khoshgoftaar, "A survey on Image Data Augmentation for Deep Learning," *J. Big Data*, vol. 6, no. 1, 2019, doi: 10.1186/s40537-019-0197-0.

[21]. Usha Ruby Dr.A, "Binary cross entropy with deep learning technique for Image classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, 2020, doi: 10.30534/ijatcse/2020/175942020.

[22] J. Bertels *et al.*, "Optimizing the Dice Score and Jaccard Index for Medical Image Segmentation: Theory and Practice," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2019, vol. 11765 LNCS, doi: 10.1007/978-3-030-32245-8_11.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:47 UTC from IEEE Xplore. Restrictions apply.

Folia Microbiologica
https://doi.org/10.1007/s12223-022-00973-6

REVIEW

Disease-associated dysbiosis and potential therapeutic role of *Akkermansia muciniphila*, a mucus degrading bacteria of gut microbiome

Vidushi Aggarwal[1] · Sushant Sunder[1] · Smita Rastogi Verma[1]

Received: 9 January 2022 / Accepted: 19 April 2022
© Institute of Microbiology, Academy of Sciences of the Czech Republic, v.v.i. 2022

Abstract
The unique functionality of *Akkermansia muciniphila* in gut microbiota indicates it to be an indispensable microbe for human welfare. The importance of *A. muciniphila* lies in its potential to convert mucin into beneficial by-products, regulate intestinal homeostasis and maintain gut barrier integrity. It is also known to competitively inhibit other mucin-degrading bacteria and improve metabolic functions and immunity responses in the host. It finds a pivotal perspective in various diseases and their treatment. It has future as a promising probiotic, disease biomarker and therapeutic agent for chronic diseases. Disease-associated dysbiosis of *A. muciniphila* in the gut microbiome makes it a potential candidate as a biomarker for some diseases and can provide future theranostics by suggesting ways of diagnosis for the patients and best treatment method based on the screening results. Manipulation of *A. muciniphila* in gut microbiome may help in developing a novel personalized therapeutic action and can be a suitable next generation medicine. However, the actual pathway governing *A. muciniphila* interaction with hosts remains to be investigated. Also, due to the limited availability of products containing *A. muciniphila*, it is not exploited to its full potential. The present review aims at highlighting the potential of *A. muciniphila* in mucin degradation, contribution towards the gut health and host immunity and management of metabolic diseases such as obesity and type 2 diabetes, and respiratory diseases such as cystic fibrosis and COVID-19.

Keywords *Akkermansia muciniphila* · Biomarker · COVID-19 · Gut microbiome · Host immunity · Mucus degradation · Obesity · Probiotic · Therapeutic · Type 2 diabetes

Introduction

Akkermansia muciniphila (*A. muciniphila*) is a recently discovered member of commensal gut microbiota and constitutes a new genus of the phylum *Verrucomicrobia* (Derrien et al. 2004). It is oval, strictly anaerobic, non-motile and Gram-negative bacteria that do not form endospores. It has circular genome of 2,664,102 base pairs, sharing 29% gene similarity with phylum *Verrucomicrobia* (van Passel et al. 2011). As unveiled by whole-genome sequencing, its proteome consists of 5644 unique proteins (Guo et al. 2017). *A. muciniphila* colonizes the gastrointestinal tracts at an early stage through human milk and accounts for 1–4% of total gut microbiota (Collado et al. 2008). The abundance of *A. muciniphila* in caecum is ubiquitous in infants and healthy adults. Besides the large intestine, it is also found in the lining of the lungs and saliva.

The importance of *A. muciniphila* lies in its potential to degrade mucin, the significant component in mucus. It consumes mucin as a carbon and nitrogen source during its life cycle and metabolism. The optimum temperature and pH for its growth are 37 °C and 6.5 respectively. A recent study showed that despite being a strict anaerobe, it could sustain lower amounts of oxygen (Ouwerkerk et al. 2016). This property is quite similar to other microbes in the intestine, especially anaerobes, such as *Bifidobacterium adolescentis* and *Bacteroides fragilis*, which can tolerate ambient amounts of oxygen for 48 h (Ouwerkerk et al. 2016). *A. muciniphila* is known to competitively inhibit other mucin-degrading bacteria and improve metabolic functions and immunity responses in the host, making it a suitable candidate as a probiotic (Belzer and de Vos 2012). *A. muciniphila*

✉ Smita Rastogi Verma
smitar@dtu.ac.in

[1] Department of Biotechnology, Delhi Technological University, Delhi 110042, India

Published online: 20 May 2022

was even found to be effective in treatment of inflammatory bowel diseases and cancer (Png et al. 2010; Chen et al. 2020a). The present review aims at studying the potential of *A. muciniphila* in mucin degradation, contribution towards the gut health and host immunity, and management of metabolic diseases such as obesity and type 2 diabetes, and respiratory diseases such as cystic fibrosis and COVID-19.

A. muciniphila in gut microbiome

The prevalence of *A. muciniphila* has been connected with a healthy gut, and therefore, its richness is inversely linked to numerous disease conditions (Jakobsson et al. 2015). An investigation of its relationships with the hosts revealed *A. muciniphila* to enhance the intestinal barrier function in mice (Shin et al. 2014). Colonization by *A. muciniphila* culminated in transcriptome alterations, leading to a rise in the genetic expression linked with immunogenicity. Outer membrane proteins of *A. muciniphila* were discovered to play a function in controlling immunological responses. One of the outer membrane proteins was recently discovered (Amuc-1100) (Ottman et al. 2017b). The work demonstrated that the outer membrane pili-like protein is essential in immunological modulation and the increase of trans-epithelial resistance. *A. muciniphila* performed a function in regulating metabolic endotoxemia and adipose tissue metabolism. Several investigations have consciously or inadvertently discovered the existence of *Akkermansia*-like spp. in regions of the human body other than the colon, where *A. muciniphila* could also have vital activities. The physiology and environmental factors of *Akkermansia*-like spp. in distinct anatomic locations of the digestive tract allow us to evaluate the ability of *A. muciniphila* to colonize and be productive at all these niches. Various beneficial effects of *A. muciniphila* in the human microbiome are presented in Fig. 1.

Mucin-degrading activity of *A. muciniphila*

Mucus consists of heavily glycosylated mucin-2 (MUC2), an oligosaccharide composed of various amino sugars and monosaccharide sugars, including *N*-acetyl-*D*-galactosamine (GalNAc), *N*-acetyl-*D*-glucosamine (GlcNAc), *D*-galactose

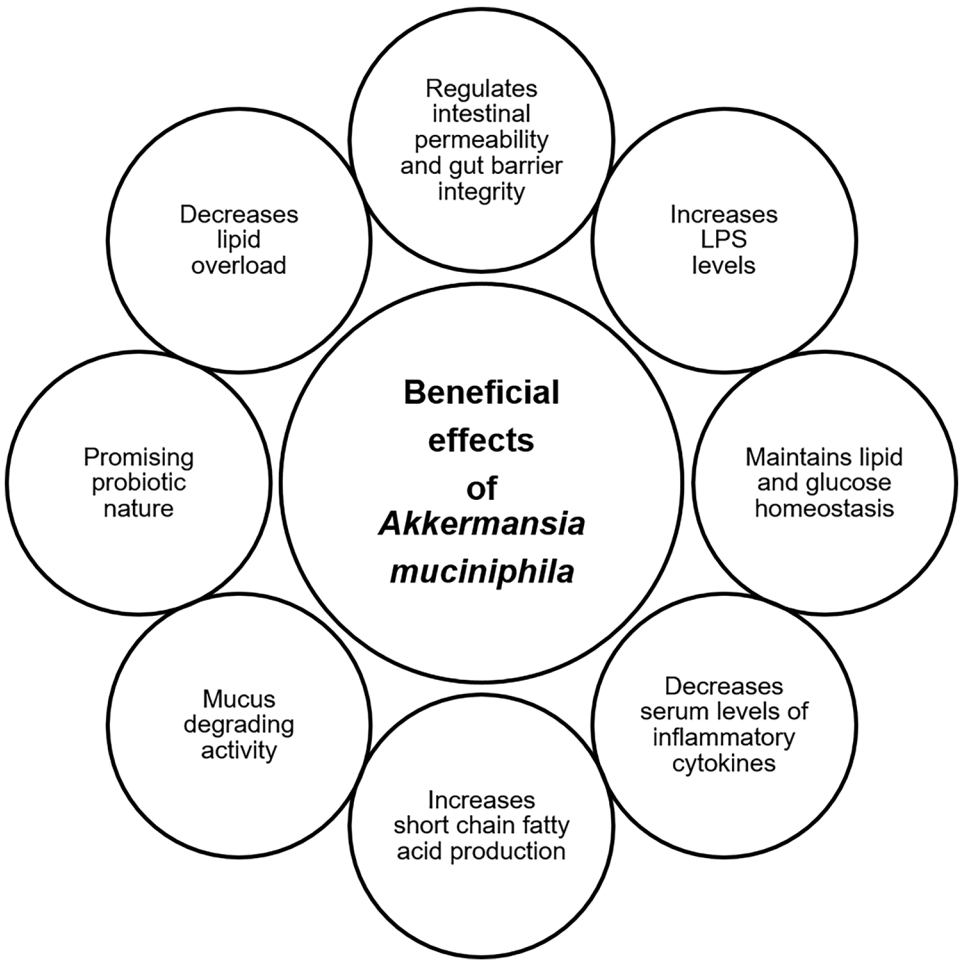


Fig. 1 Beneficial effects of *A. muciniphila* in human microbiome

and *L*-fucose (Ottman et al. 2017a). In many cases, these sugars are further substituted with acetate, phosphate and sulfate groups. Mucin has defining roles like a lubricant for food transport over membranes and provides selective permeability that allows the flow of nutrients to epithelial cells. It also acts as the first line of defence against mechanical damage, pathogens, and toxins and provides a surface layer to bacteria for its growth, adhesion and protection (Cone 2009; Johansson et al. 2013). However, some bacterial species of human microbiota release inflammatory toxins, which increase the permeability of the mucus layer and ultimately decrease its barrier property (Jakobsson et al. 2015). It has been concluded that bacterial colonies reside only at the outer layer of the intestinal tract. In contrast, the inner layer intends to keep the bacteria at bay from the epithelial cells to enforce immune tolerance to the guts by transporting IgA and antimicrobial proteins (Johansson et al. 2008, 2011).

It is observed that *A. muciniphila* can maintain an exciting microbial relationship in the host intestine by converting mucin into beneficial by-products (Ottman et al. 2017a, b). Recent studies showed that *A. muciniphila* could be grown on a synthetic medium where mucin can be replaced with media containing glucose, threonine, peptone and GlcNAc (Plovier et al. 2016). The amino group of sugars promotes the growth of bacteria in the presence of casitone, tryptone, yeast extract and peptone. One of the essential factors that account for the proliferation of *A. muciniphila* is glucose-6-phosphate, one of the constituents of mucin known to promote the adaptation of mucosal niche (van der Ark et al. 2018). In order to study substrate uptake abilities of *A. muciniphila*, few studies were conducted on a genome-scale metabolic model to demonstrate amino acids auxotrophy, sugar degrading capacities and vitamin biosynthesis (Ottman et al. 2017a). These experiments have also been validated through in vivo experiments in which *A. muciniphila* has been shown to proficiently utilize mucin-derived monosaccharide sugars and amino sugars. It has been found that the uptake of mucin-derived sugars and non-mucin sugar glucose by *A. muciniphila* is enhanced in a mucin-rich environment which indicates the need of mucin-derived components for the optimal growth of bacteria. In vivo experiments have also suggested that *A. muciniphila* may have galactose metabolism; however, mucin-derived components are necessary for its growth.

Transcriptomic analysis of *A. muciniphila* under mucin-rich and mucin-depleted conditions showed differential gene expression suggesting a global change in cellular functions (Shin et al. 2019). Out of 1126 differentially expressed genes (DEGs), 583 genes were upregulated while 543 were downregulated in mucin-rich conditions as compared to mucin-depleted conditions. The upregulated genes were significantly related to hydrolase activity acting on glycosidic bonds and their transporters, thereby confirming the activity of mucin-degrading genes under mucin-rich conditions. Thus, the genes that encode mucin-degrading enzymes, such as sulfatases, galactosidases, acetyl-glucosaminidase, neuraminidases and L-fucosidase transporters were upregulated in mucin-rich conditions. Furthermore, their downregulation in mucin-depleted medium determined the importance of its role in mucin-degradation. The catabolic glycolysis pathway is also correlated with mucin-degradation pathways (Shin et al. 2019). Thus, under mucin-depleted conditions, genes involved in glycolysis and energy metabolism, such as NADH dehydrogenase, succinate dehydrogenase and ATP synthase, either showed similar expression levels or were upregulated significantly. At the same time, there were few exceptions, including one ATP-dependent 6-phosphofructokinase gene (Amuc_1481), two enolase genes (Amuc_844, Amuc_1184) and one dihydrolipoyl dehydrogenase gene (Amuc_1689).

A. muciniphila in host immunity and probiotic nature

The symbiotic relationship between the gut microbiota and host determines the normal physiology, immunity and pathogen susceptibility of an individual. The interplay between host and gut microbiome that helps in pathogen displacement, regulating immune response and anti-inflammatory pathways, is a vital phenomenon. There is abundant evidence in the literature on mucin-utilizing *A. muciniphila* conferring immunity (Tummler and Puchelle 1997; Plovier et al. 2016; Ottman et al. 2017a). Many mucin degradation pathways regulate the host pathway by signalling through tumour necrosis factor α (TNF-α), interferon γ, interleukins-10 (IL-10) and interleukins-4 (IL-4) (Derrien et al. 2011; Andersson et al. 2012; Collado et al. 2012). Decreased levels of anti-inflammatory cytokines (IL-10 and IL-4) induced interleukins, while increased proinflammatory cytokines (TNF-α and IFN-γ) causing rapid proliferation of *A. muciniphila*. An increase in 2-arachidonoylglycerol levels was noted post *A. muciniphila* treatment, reducing inflammation (Gunderson and Kopito 1994; Everard et al. 2013). The secreted proteins from bacteria interact with host immune cells to induce signalling pathways that exhibit anti-inflammatory and immunomodulatory activity (Sánchez et al. 2008, 2010; Bernardo et al. 2012; Ruiz et al. 2014). The extracellular material secreted from it activates the downstream signalling pathway like toll-like receptors 2 (TLR2) (Ottman et al. 2017b). Amuc_110, a specific protein in the outer membrane, recapitulates the effect of bacteria on TLR2 activation and improves the barrier integrity of intestines (Dean and Annilo 2005; Belzer and de Vos 2012; Plovier et al. 2016; Ottman et al. 2017b). However, it is still unknown how Amuc_110 protein is regulated in the presence of a dynamic mucosal environment. The gene encoding

Amuc_110 protein is highly regulated in a mucin-depleted environment (Plovier et al. 2016). Amuc_110 has also shown its ability to exert a probiotic effect on diet-induced obesity and was present in the extracellular proteins of *A. muciniphila* as well (Plovier et al. 2016). To conclude, *A. muciniphila* is inversely correlated with inflammatory conditions and helps in epithelial barrier integrity by stimulating anti-inflammatory pathways (Gunderson and Kopito 1994; Shin et al. 2014; Cantarel et al. 2015; Caesar et al. 2015; Schneeberger et al. 2015).

The composition and functioning of the human gut microbiota are primarily proportional to nutritional accessibility of microbiota either obtained from a host or food (Zoetendal et al. 2012; Nicholson et al. 2012; Salonen and de Vos 2014). *A. muciniphila* is one of the good bacteria of the human gut microbiota. The presence of *A. muciniphila* in the intestinal mucus layer indicates it to be involved in gut regulation and host metabolism. It exists in a symbiotic relationship with the mucosal layer, and its abundance is greatly affected by the nutrients present in the mucus layer located around the intestinal epithelial cells. Its presence also supports other beneficial bacteria in the gut microbiome. *A. muciniphila* catabolizes mucins and turns them into short-chain fatty acids (SCFAs), including acetate, which other beneficial bacteria exploit, such as *Firmicutes*, to produce butyrate, a vital source of energy for the cells lining the gut. The production of SCFAs from the breakdown of mucin supplies energy to the goblet cells, which are responsible for secreting mucins. Furthermore, the consumption of specific dietary fibres can increase the abundance of this friendly bacteria, which helps thicken the mucus lining the gut. This strengthens the gut lining and improves gut barrier function and may, in turn, help in preventing weight gain. Chelakkot et al. demonstrated the role of Amuc_1100, isolated from *A. muciniphila*, in AMP-activated protein kinase (AMPK) activation mechanism, thereby improving gut integrity (Chelakkot et al. 2018). Amuc_1100 has been implicated in enhancing the expression of tight junction protein-1 (Tjp-1) and occludin (Li et al. 2016), thereby contributing to the gut barrier function. Thus, the presence of *A. muciniphila* in the mucus of the intestine regulates intestinal homeostasis and its integrity barrier through host signalling pathways (Derrien et al. 2004; Ottman et al. 2017a).

A. muciniphila dysbiosis associated with disease states and its management

The gut microbiome of healthy people is quite diverse. Gut microflora through microbial antigens and metabolites is the master regulator of innate and adaptive immunity. Disease associated dysbiosis results in induction, training and function of immune system. The disturbance of the mucus layer by any means may lead to inflammation and increase the risk of infection. Even slight variance in intestinal flora may be associated with sensitivity and severity of a disease. Therefore, gut microbes find use as potential diagnostic biomarkers by studying their abundance in different diseases.

A. muciniphila has been linked with wide range of diseases and disorders such as type 2 diabetes (Tilg and Moschen 2014), alcoholic steatohepatitis (ASH) (Grander et al. 2018), appendicitis (Swidsinski et al. 2011), obesity (Dao et al. 2016), atopic diseases (Drell et al. 2015), colorectal cancer (Weir et al. 2013), autism (Wang et al. 2011), inflammatory bowel disease (Png et al. 2010), cystic fibrosis (Hayden et al. 2019), and COVID-19 (Yeoh et al. 2021) (Table 1). Various studies demonstrated *A. muciniphila* to be negatively correlated with inflammatory bowel diseases (Png et al. 2010; Rajilić-Stojanović et al. 2013), appendicitis (Swidsinski et al. 2011), obesity (Karlsson et al. 2012; Dao et al. 2016) and type 2 diabetes (Tilg and Moschen 2014). Lower abundance of *A. muciniphila* is commonly observed in the majority of metabolic disorders while higher abundance is seen in few cases like colorectal cancer (Yu et al. 2017).

Recently, an association between *A. muciniphila* and metastasis of lymph nodes in lung adenocarcinoma was reported (Chen et al. 2020a). The relative abundance of *A. muciniphila* was observed to be greater in metastasis cohort (0.057) than the non-metastasis cohort (0.023) pointing towards its potential as a promising biomarker to predict lymph node metastasis. Likewise, a study to investigate whether *A. muciniphila* could enhance the antitumor effect of cisplatin (*cis*-diamminedichloroplatinum; CDDP) was conducted (Chen et al. 2020b). It was found that when *A. muciniphila* was combined with cisplatin, the growth of tumour volume slowed and the changes of tumour pathomorphology significantly improved. At molecular level, upregulation of factor-associated suicide (Fas) proteins and downregulation of p53, ki-67 and Fas ligand (FasL) proteins were also observed. Several proinflammatory factors (TNF-α, IFN-γ and IL-6) were induced while the expression of $CD4^+CD25^+Foxp3^+$ Treg was suppressed indicating the role of *A. muciniphila* in regulating immune inflammatory microenvironment in reversion of tumour growth and tumour immune escape. Also, the levels of IFI27l2 and IGFBP7, two most differentially expressed genes in lung cancer, were found to be increased because of the combined treatment with *A. muciniphila* and CDDP. Signalling pathways such as JAK-STAT, FOXO, cytokine-cytokine receptor interaction, PI3K-Akt, Th17 and cell differentiation were associated with the antitumor effect of *A. muciniphila* and CDDP. The study suggested that *A. muciniphila* combined with CDDP provides good symbiotic environment to achieve maximum therapeutic efficacy of antitumor drugs. The human gut microbiome could

Table 1 Altered abundance of *A. muciniphila* in various disease states in humans

Disease	*A. muciniphila* abundance	Detection method	Sample type	References
Allergic asthma	Reduced	qPCR	Faeces	(Demirci et al. 2019)
Asthma	Reduced	16S rRNA sequencing	Faeces	(Michalovich et al. 2019)
Alcoholic steatohepatitis (ASH)	Reduced	16S rRNA sequencing	Faeces	(Grander et al. 2018)
Atopy	Reduced	16S rRNA sequencing	Faeces	(Candela et al. 2012)
Atopy	Reduced	Pyrosequencing	Faeces	(Drell et al. 2015)
Autism	Elevated	bTEFAP	Faeces	(De Angelis et al. 2013)
Autism	Reduced	qPCR	Faeces	(Wang et al. 2011)
Clostridium difficile infection	Elevated	qPCR	Faeces	(Vakili et al. 2020)
Colorectal cancer	Elevated	16S rRNA sequencing	Faeces	(Weir et al. 2013)
Colorectal cancer	Elevated	qPCR	Tissue biopsy	(Mira-Pascual et al. 2015)
Crohn's disease	Reduced	qPCR	Tissue biopsy	(Png et al. 2010)
Crohn's disease	Reduced	16S rDNA pyrosequencing	Faeces	(Opstelten et al. 2016; Malham et al. 2019)
Cystic fibrosis	Reduced	16S-based tag-encoded FLX amplicon pyrosequencing (bTEFAP)	Faeces	(Hoffman et al. 2014)
Cystic fibrosis	Reduced	Metagenomic sequencing	Faeces	(Hayden et al. 2019)
COVID-19	Elevated	Shotgun sequencing	Faeces	(Yeoh et al. 2021)
Oesophageal cancer	Elevated	16S rRNA sequencing	Tissue biopsy	(Snider et al. 2019)
Hyperlipidaemia	Reduced	16S rRNA sequencing	Faeces	(Gargari et al. 2018)
Microscopic colitis	Reduced	16S rDNA pyrosequencing	Faeces	(Fischer et al. 2015)
Multiple system atrophy	Elevated	Metagenomic sequencing	Faeces	(Wan et al. 2019)
Obesity	Reduced	qPCR	Faeces	(Marvasti et al. 2020)
Obesity	Elevated	qPCR	Faeces	(Remely et al. 2015)
Parkinson's disease	Elevated	qPCR	Faeces	(Unger et al. 2016)
Parkinson's disease	Elevated	Shotgun sequencing	Faeces	(Bedarf et al. 2017)
Prediabetes	Reduced	16S rRNA sequencing	Faeces	(Allin et al. 2018)
Psoriasis	Reduced	16S rDNA pyrosequencing	Faeces	(Tan et al. 2018)
Pulmonary arterial hypertension	Reduced	Shotgun sequencing	Faeces	(Kim et al. 2020)
Pulmonary tuberculosis	Reduced	Metagenomic sequencing	Faeces	(Hu et al. 2019)
Schizophrenia	Elevated	16S rRNA sequencing	Faeces	(Xu et al. 2020a)
Schizophrenia	Elevated	Shotgun sequencing	Faeces	(Zhu et al. 2020)
Spleen deficiency syndrome	Reduced	qPCR	Faeces	(Peng et al. 2020)
Type 1 diabetes	Reduced	qPCR	Faeces	(Fassatoui et al. 2019)
Type 2 diabetes	Elevated	Metagenomic sequencing	Faeces	(Chelakkot et al. 2018)
Type 2 diabetes	Elevated	Shotgun sequencing	Faeces	(Qin et al. 2012)
Type 2 diabetes	Reduced	16S rRNA sequencing	Urine	(Liu et al. 2017)
Type 2 diabetes	Reduced	Metagenomic sequencing	Faeces	(Zhong et al. 2019)
Type 2 diabetes	Reduced	qPCR	Faeces	(Fassatoui et al. 2019)
Ulcerative colitis	Reduced	qPCR	Tissue biopsy	(Png et al. 2010)
Ulcerative colitis	Reduced	MiSeq sequencing	Faeces	(Bajer et al. 2017)
Ulcerative colitis	Reduced	16S rRNA sequencing	Faeces	(Malham et al. 2019)

therefore, be responsible for differences in drug response of individuals paving way for personalized therapeutics to significantly improve human health care. Thus, by regulating the abundance of *A. muciniphila* in the gut microbiome in a personalized manner, early treatment of related diseases may be facilitated. The microbiome in general, and *A. muciniphila* in particular, can act as a biomarker of these diseased states and provide future theranostics by suggesting ways of disease diagnosis and treatment (Morgan and Huttenhower 2012).

A. muciniphila and metabolic diseases

Decreased abundance of *A. muciniphila* in obesity

World Health Organization (WHO) defines obesity as abnormal fat accumulation leading to health implications like type 2 diabetes, fatty liver disease and hypertension. Being a multifactorial disorder, not only does it result in fatal complications like cardiovascular diseases and psychological effects but also challenges one to perform everyday tasks with ease. It is a low-grade inflammatory metabolic disorder resulting in higher levels of inflammatory cytokines such as IL-6, TNF-α and hypersensitive C-reactive protein (hs-CRP). Inflammation in obesity leads to significant changes in gut microbiota, for example, increase in the abundance of *Firmicutes, Bifidobacterium spp.* and *Lactobacillus gasseri* while decrease in the abundance of *Bacteroidetes* (Ley et al. 2005; Wang and Jia 2016). *A. muciniphila* was inversely associated with obesity and found to be more abundant in lean individuals than overweight individuals (Remely et al. 2016). The various mechanisms by which *A. muciniphila* helps in obesity management are shown in Fig. 2. *A. muciniphila* maintains the intestinal immunity, gut barrier integrity and permeability by reducing inflammatory cytokines, thereby achieving metabolic homeostasis (Ottman et al. 2017b). Also, lipopolysaccharide (LPS) found in cell wall of Gram-negative bacteria is a potential proinflammatory molecule and is associated with onset of inflammation. Imbalance in LPS levels leads to activation of proinflammatory signalling pathways causing increased secretion of IL-6 and TNF-α. The inherent property of *A. muciniphila* of SCFAs production signals G-protein coupled receptor (GPCR) activation and histone deacetylase (HDAC) inhibition to maintain energy homeostasis and appetite sensation (Lukovac et al. 2014). SCFAs production facilitate increased production of glucagon-like peptide 1 (GLP-1) and glucose-dependent insulin tropic polypeptide (GIP) upon binding to GPR41/GPR43 receptors present in L-cells in the intestinal mucosa. This results in improvement in insulin resistance and glucose tolerance thereby, suppressing appetite through metabolic signalling. Furthermore, *A. muciniphila* decreases inflammation by increasing the levels of oleoylathanolamide (OEA), 2-palmitoylglycerol (2-PG), 2-acylglycerol (2-AG) and 2-oleoylglycerol (2-OG) which bind to GPR119 receptors, stimulating release of GLP-1 (Everard et al. 2013). In a recent study, oral supplementation of *A. muciniphila* in obese mice through high fat diet (HFD) has been demonstrated to reduce the intestinal endotoxin levels, hence reducing inflammation (Cani and de Vos 2017; Fuke et al. 2019). It helped restore gut barrier dysfunction through symbiotic relationships with other beneficial microbes like *Bacteroidetes*, *Euryarchaeota*, *Firmicutes* and *Actinobacteria* and improved intestinal permeability through the inhibition of claudin 3 (Cldn3), cannabinoid receptor 1 (Cnr1) and occludin like tight junction proteins or lowering of flavin-containing monooxygenase 3 (FMO3) expression (Dao et al. 2016).

In humans, both live and pasteurized *A. muciniphila* were found to be safe for oral consumption by heavy body weight individuals (Plovier et al. 2016). In a clinical trial (NCT02637115), oral administration of *A. muciniphila* in obese patients for 3 months was found to be an effective and safe treatment. Similarly, in a randomized, double-blind human study based on *A. muciniphila* supplementation, a decrease in body weight along with improvement in liver dysfunction and inflammation in patients was observed (Depommier et al. 2019). Also, supplementation with prebiotic containing oligofructose helped restore *A. muciniphila* abundance. But since *A. muciniphila* does not grow in vitro

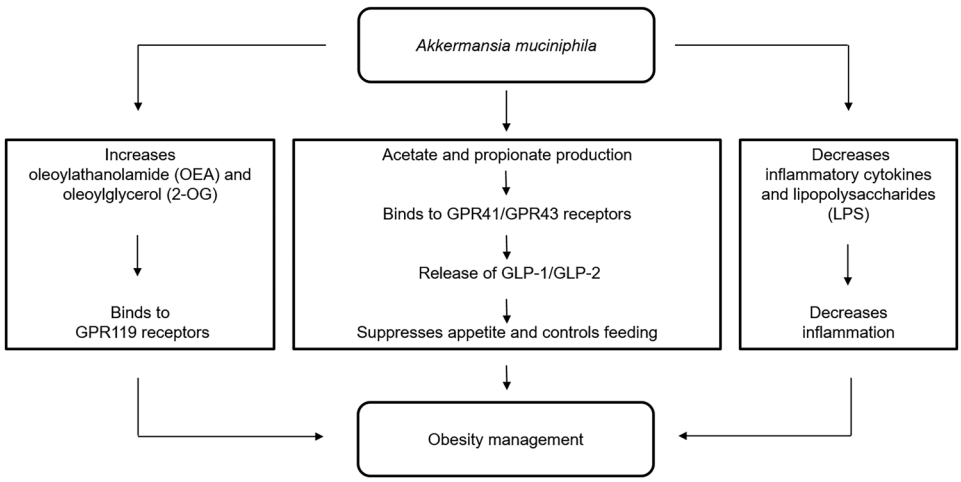


Fig. 2 Role of *A. muciniphila* in obesity management

on oligofructose-enriched media, it can be concluded that complex cross-feeding interactions might be involved. While it is clear that the human mucus colonizer maintains gut barrier integrity and homeostasis during obesity, the role of human gut microbiome on etiology of obesity remains to be investigated. The study of *Akkermansia*-obesity relationship could provide better insights to microbe-based treatments. Since *A. muciniphila* regulates the energy metabolism of host, its therapeutic intervention in obesity could be explored (Xu et al. 2020b). The reduction in fat-mass ratio in obesity can be studied by urinary metabolomics profile of *A. muciniphila*, making it a suitable biomarker for obesity (Png et al. 2010). Furthermore, the risk of obesity and associated metabolic diseases can be reduced by proper management of gut microbial profile.

Decreased abundance of *A. muciniphila* in type 2 diabetes

According to WHO, 1 out of 11 people suffer from diabetes with over 1.5 million deaths globally. Type 2 diabetes is a silent killer marked by the body's inability to utilize insulin produced by pancreatic β-cells. A study conducted by National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) suggests that factors such as sedentary lifestyle and genetic conditions may contribute to early onset of type 2 diabetes. The increased levels of blood sugar, i.e. hyperglycaemia results in nervous, circulatory and immune system–related complications. Insulin resistance as a result of type 2 diabetes affects the gut microbial diversity and metabolite production. Alterations in the abundance of various gut microbes during the onset and progression of type 2 diabetes have also been observed. Many studies have reported negative correlation of genus *Akkermansia, Bacteroides, Faecalibacterium, Roseburia* and *Bifidobacterium*, while positive correlation of genus *Ruminococcus*, *Blautia* and *Fusobacterium* with type 2 diabetes (Sedighi et al. 2017; Gao et al. 2018). Patients with normal glucose tolerance exhibited higher abundance of *A. muciniphila* as compared to pre-diabetic or type 2 diabetes patients. Interestingly, several anti-diabetic drugs like metformin, dapagliflozin and liraglutide were found to favour the abundance of *A. muciniphila* (Shin et al. 2014; Wang et al. 2018; Lee et al. 2018). Furthermore, a study conducted on the administration of an antidiabetic drug, metformin revealed that *A. muciniphila* further enhanced its anti-diabetic effects (Shin et al. 2014). Mice fed on HFD, when treated with metformin, showed increased abundance of *A. muciniphila* and improved blood sugar levels. Moreover, improved tolerance to glucose was observed upon oral administration of *A. muciniphila* but not metformin.

A. muciniphila is known to protect the intestinal barrier function by maintaining normal blood sugar levels. The intrinsic ability of *A. muciniphila* to catabolize complex carbohydrates further assists in inhibition of α-glucosidase and reduction of postprandial hyperglycaemia (Everard et al. 2013). Both obesity and type 2 diabetes may be ameliorated by increasing fatty acid oxidation and energy expenditure but reducing fatty acid biosynthesis (Everard et al. 2013; Gurung et al. 2020). Thus, fatty acid oxidation in adipose tissues and adipocyte differentiation may be promoted by administration of *A. muciniphila*, and other bacteria such as *Lactobacillus gasseri*, *Bacteroides acidifaciens* and thus SCFAs. This is correlated to increased levels of 2-PG, 2-AG and 2-OG in the adipose tissue. Furthermore, *A. muciniphila* can be regulated by increasing circulation of tryptophan metabolites through dietary intake (Cronin et al. 2021). Interestingly, *A. muciniphila* positively influences the host's glucose metabolism by inducing IL-10, thus protecting from ageing-related insulin resistance (Wang et al. 2015; Greer et al. 2016). It fights against diabetic oxidative damage and improves resistance to gluco/lipotoxicity by decreasing hepatic glycogen levels and increasing HDL-C levels (Zhang et al. 2018). Moreover, it reduces inflammation by inhibiting the expression of TNF-α and lipid oxidative damage by lowering malondialdehyde levels in diabetic animals (Zhang et al. 2018). Increased paracellular gut permeability and gut barrier disruptions result in inflammation and metabolic diseases by increased absorption of LPS. *A. muciniphila* maintains glucose homeostasis and fat mass storage by controlling host mucus turnover and higher L-cell activity. Therefore, prebiotic treatment could help restore *A. muciniphila* abundance and counter metabolic endotoxemia in type 2 diabetes and obesity. Therefore, it becomes necessary to study the host-gut microbe interactions for better understanding of governing mechanisms in development of type 2 diabetes. *A. muciniphila*, through the alteration of the gut microbiota, could help to control and manage type 2 diabetes in the near future.

A. muciniphila and respiratory diseases

Decreased abundance of *A. muciniphila* in cystic fibrosis

Cystic fibrosis (CF) is an autosomal recessive disorder caused by mutations in the cystic fibrosis transmembrane conductance regulator (CFTR) gene located on chromosome 7. CFTR protein functions as an important anion-selective ion channel responsible for epithelial fluid secretion and intra-luminal hydration (Welsh and Smith 1993). Defective CFTR protein leads to accumulation of mucus in the lungs and intestine largely affecting the pulmonary and intestinal microbiota (Price and O'Toole 2021). As the mucus gets more and more viscous, the mucociliary clearance mechanism (MCC) becomes unable to clear the mucus resulting in manifestation of opportunistic microbial infections. Dysfunctional CFTR results in an altered intestinal condition including dysbiosis of gut

microbiota due to physiological and biochemical imbalance. It includes changes in intestinal pH, inflammation, malabsorption and gut barrier disruptions (Meeker et al. 2020). Various factors such as prolonged antibiotics intake, immunosuppressive mediations and high-calorie diet further shape the CFTR microbiome. Specifically, in delF508 mutations, bacteria such as *E. coli* and *Eubacterium biforme* were found in higher abundance while *Bifidobacterium* and *Faecalibacterium* species were in lower abundance (Schippa et al. 2013). Recent studies on the gut microbiota of CFTR-deficient mice reported an increase in abundance of *Enterobacteriaceae*, *Mycobacteria* and *Bacteroides* while a decrease in abundance of *Lactobacilli*, *Acinetobacter lwoffii* and *A. muciniphila* (Thomsson et al. 2002; Bazett et al. 2016). Through integrated metagenomics and metabolomics study on CFTR gut microbiota, an increase in the expression of associated metabolites such as propylbutyrate, γ-aminobutyric acid (GABA), ethanol, choline and pyridine was observed while there was reduction in expression levels of 4-methylphenol, methylacetate, uracil, sarcosine, acetate, phenol, benzaldehyde and glucose (Vernocchi et al. 2018). The multi-omics-based model pointed out the correlation of microbial and metabolite variations caused by CFTR functional defects. Therefore, it becomes important to study the relationship between host gut, microbes and associated metabolites to investigate CF and its biomarkers. Major evidence of their relationship is the study on the administration of CFTR potentiator Ivacaftor which resulted in an increase in *A. muciniphila*. It could be explained by the release of bicarbonates from CFTR post Ivacaftor treatment that provided the optimal environment for mucin degraders (Ambort et al. 2012; Schütte et al. 2014). Post treatment with Ivacaftor, a significant reduction in stool calprotectin, a protein released by neutrophils, but no change in M2 pyruvate kinase (M2-PK) was observed. The reduction of calprotectin indicated that intestinal inflammation could be improved in CF patients upon restoration of intestinal milieu. Also, there was selective loss of pathogens like those of *Enterobacteriaceae* family which was positively correlated with stool calprotectin level (Manor et al. 2016). It led to an increase in the expression of an antimicrobial peptide (Reg III), which has direct metabolic activity in the intestine against Gram positive bacteria. In CF, *A. muciniphila* accounted for normal stool M2-PK concentration and decreased amount of *Enterobacter.* Thus, the increased abundance of *A. muciniphila* supports its potential as a biomarker for the gut and it may be used for the microbe based therapy in CF (Pang et al. 2014). Many gut microbes including *A. muciniphila* could be associated with CF and hold future as its promising therapeutic intervention. By regulating the gut profile of *A. muciniphila* through personalized nutrition and supplementation, host immunity can be improved, which could serve as one of the prophylactic ways by which the severity of CF could be minimized.

Increased abundance of *A. muciniphila* in COVID-19

On 11th March 2020, COVID-19 caused by severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) was declared as a pandemic. Over 180 countries were affected globally leading to nationwide lockdowns. The virus primarily attacked the respiratory system causing high-grade fever, severe cough, shortness of breath and pneumonia in severe cases. Several people infected with the virus experienced neurological and gastrointestinal (GI) manifestations with or without respiratory symptoms. On the other hand, some people were asymptomatic or symptom-free. The viral infection was also correlated with gut-lung-brain axis and microbiome imbalance. Significant reduction in the abundance of beneficial microbes was associated with inflammation and pathogenesis in COVID-19 (Hussain et al. 2021). Altered microbiome composition could further weaken body's immunity and may play a role in SARS-CoV-2 infection. Recent studies have elucidated the relationship between gut and lung microbiota in COVID-19 and potential as prognostic markers (Wang et al. 2021). Although there is no direct evidence of specific interaction between resident gut microbe and COVID-19, some studies suggest that the gut microbiome in COVID-19 could be a key player in modulating host response and disease severity (Hussain et al. 2021; Yamamoto et al. 2021; Yeoh et al. 2021). GI symptoms were accompanied by gut dysbiosis during the early phase causing changes in gut microbiome and increase in inflammatory cytokines. Recently, proinflammatory cytokine storm due to significant increase in levels of IL-6 and IL-10 was reported to be predictive of COVID-19 severity (Han et al. 2020). Moreover, the presence of SARS-CoV-2 RNA was reported in faecal samples suggesting gut to be a viral replication site (Xiao et al. 2020). According to a study performed on SARS-CoV-2 recovered patients, *A. muciniphila* along with *B. dorei* was found to be elevated in the COVID-19 patients (Yeoh et al. 2021). Moreover, these bacteria were positively correlated with inflammatory cytokines, namely IL-1β and IL-6 and proinflammatory cytokine C-X-C motif ligand 8 (CXCL8). On the other hand, *Faecalibacterium prausnitzii*, *Eubacterium rectale* and some species of *Bifidobacteria* were found in lower abundance. A recent faecal metabolomics studies through machine learning approach suggested that particular set of gut microbiota could be used to predict proteomic risk score based on 20 blood proteomic biomarkers for COVID-19 severity (Gou et al. 2021). The gut microbial profile was used as a tool for prediction of the blood molecular signatures, indicating amino acid related pathways such as aminoacyl-tRNA biosynthesis, arginine biosynthesis

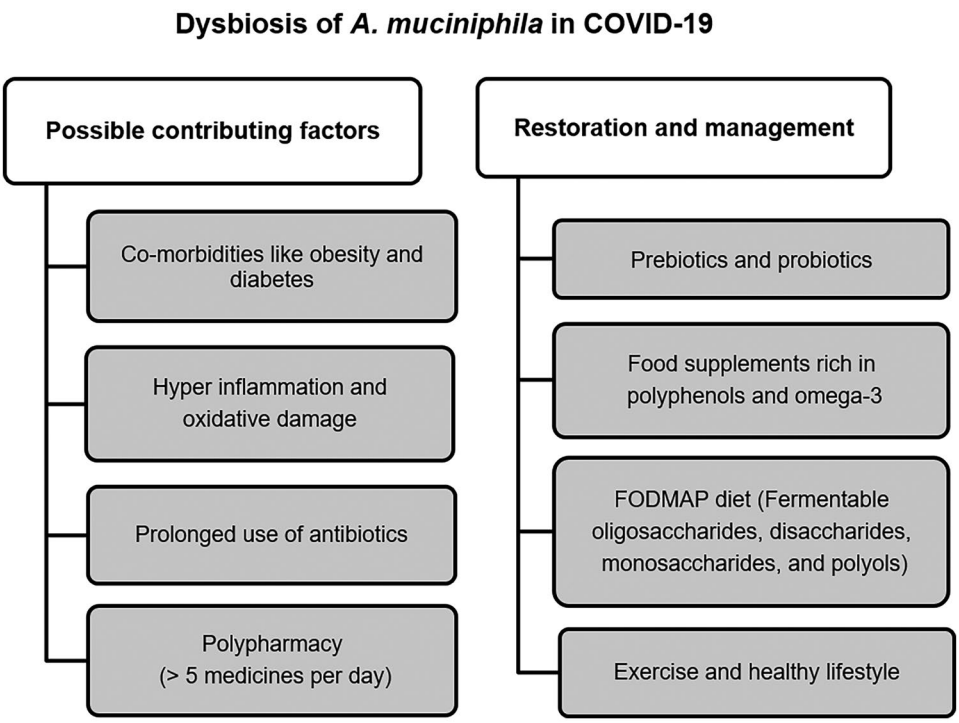


Fig. 3 Possible factors contributing to *A. muciniphila* dysbiosis in COVID-19 along with strategies to restore its normal abundance

and valine, leucine and isoleucine biosynthesis to be a possible link between inflammation and gut microbiota. It is well known that the gut microbiome renders beneficial effects on pulmonary mucosal immunity and host defence, thus safeguarding against respiratory infections (Gray et al. 2017). Downregulation of angiotensin-converting enzyme II (ACE2) involved in amino acid transport, tryptophan and antimicrobial peptide metabolism upon binding with viral spike protein might affect gut microbial ecology leading to dysbiosis in COVID-19 (Kuba et al. 2005). The possible factors contributing to *A. muciniphila* dysbiosis in COVID-19 along with strategies to restore its normal abundance are shown in Fig. 3. Co-morbidities such as diabetes, obesity and cardiovascular diseases largely influence the risk of infection and severity of disease. A lot of stress and consumption of fat and carbohydrate rich foods during the quarantine period was observed, leading to reduced CD8 + T cell response which could be linked to higher risk of infection (Mattioli et al. 2020). On the contrary, non-pharmacological measure such as reduction in consumption of fast food and increased emphasis on a healthy balanced diet also helped mitigate severe health conditions. Lifestyle habits, including diet and physical exercise, can profoundly influence the composition of the microbiome and consequently host metabolism and well-being. With the regulation of dietary habits, the optimum abundance of this friendly bacteria can be achieved in COVID-19 (Dhar and Mohanty 2020). Food supplements rich in polyphenols, omega-3 and FODMAP (Fermentable oligosaccharides, disaccharides, monosaccharides and polyols) are well known to increase *A. muciniphila*. Thus, by consuming polyphenol-rich foods, like fruits and vegetables, the abundance of *A. muciniphila* in the gut can be enhanced. Probiotic supplementations along with standard therapies as a prophylactic measure could also help mitigate the increased risk of comorbidities and move towards effective treatment. It becomes necessary to manage gut microbiome during and post disease recovery. The gut microbiome is asserted to critically impact the severity of infection as well as host immunity in COVID-19. By monitoring the GI symptoms, early diagnosis and treatment of COVID-19 can be facilitated. Therefore, it becomes important to study interaction between coronavirus and intestinal microbiome to develop novel treatment approaches.

Conclusion and future perspectives

A. muciniphila is a key mucus degrading bacteria in host immunity and infection response. The correlated metabolites and pathways are closely associated with inflammation, post-infection severity and recovery. *A. muciniphila* fortifying the intestinal mucus layer promotes several health-mediating effects. It regulates intestinal homeostasis and helps in maintaining epithelial barrier integrity by stimulating anti-inflammatory pathways. *A. muciniphila* presents itself as a powerful gut microbe having many metabolic interventions and as a promising therapeutic agent. The close relation of intestinal anti-inflammatory and protective effects of *A. muciniphila* emphasizes on its promising probiotic role (Neef and Sanz 2013). On the basis of

cross-talk elucidated across gut-lung axis, alterations in the gut microbiota through administration of SCFAs, probiotics or micronutrients could act as potential therapeutic strategies. A specific protein in the outer membrane of *A. muciniphila*, Amuc-110, which recapitulates the effect of bacteria on TLR2 activation and improves the barrier integrity of intestines, could serve as a strong candidate for drug production in future. Thus, by unveiling the interrelationships between host factors such as diet, lifestyle habits, clinical markers and *A. muciniphila* in the human gut microbiome, interventions and clinical trials may be designed. Through extensive investigation, new dimensions of the impact of *A. muciniphila* in the microbiome on human health may be explored in obesity, type 2 diabetes, cystic fibrosis and COVID-19 by examining its dysbiosis in patients as compared to healthy individuals and adopting suitable strategies for its restoration (Fig. 4). The study of the relationship between *A. muciniphila* in gut microbiome and personalized medicine can be one of the most attractive aspects of future research, which can provide significant perspectives for the treatment of metabolic diseases like type 2 diabetes and obesity, and even respiratory diseases such as cystic fibrosis and COVID-19. To conclude, potential opportunities exist for targeted interventions to modify the composition of *A. muciniphila* in the gut microbiome to improve host health.

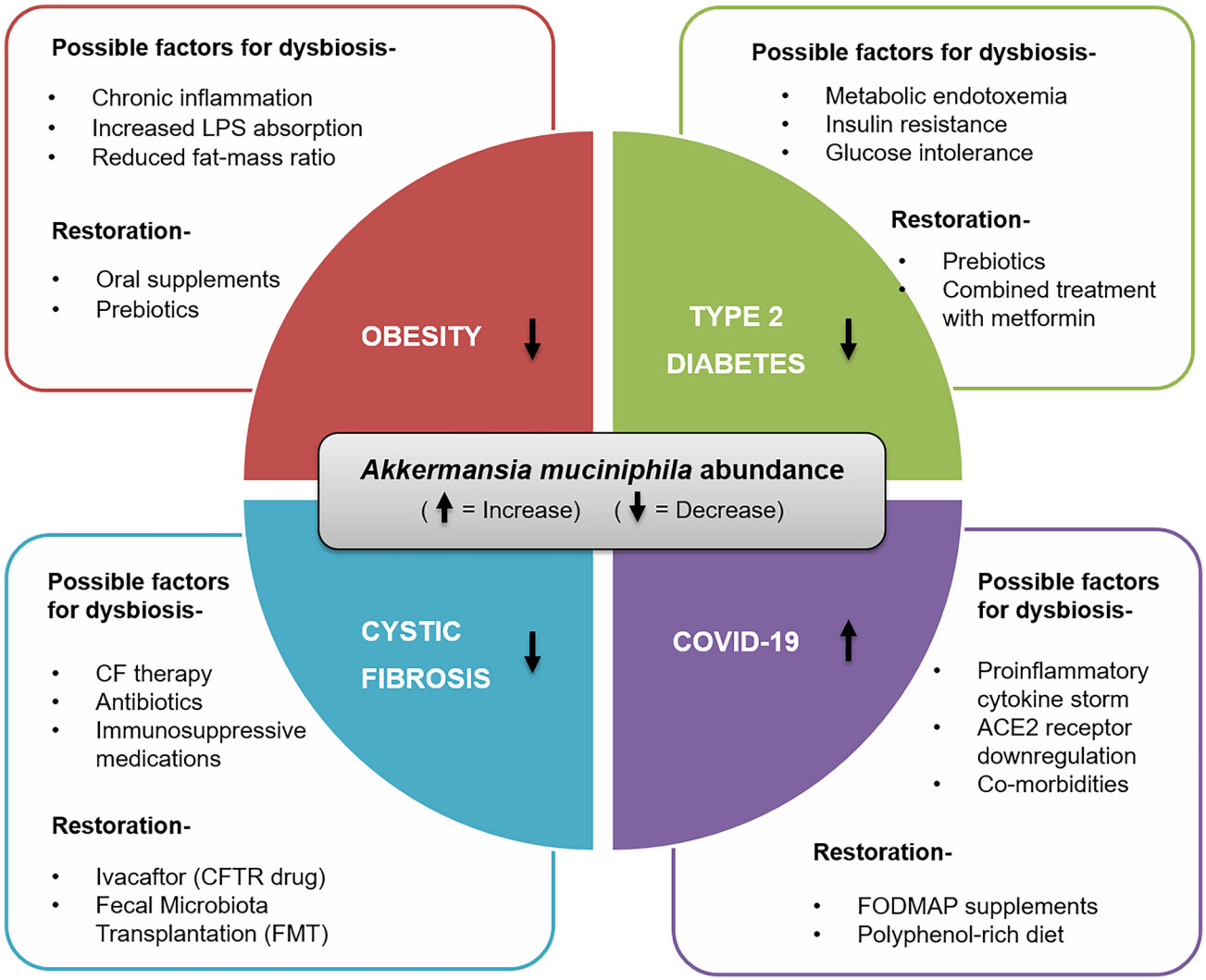


Fig. 4 Dysbiosis and restoration of *A. muciniphila* in obesity, type 2 diabetes, cystic fibrosis and COVID-19

Acknowledgements The authors are grateful to the Delhi Technological University, New Delhi, India, for providing support to conduct this study.

Author contribution Conceptualization: Vidushi Aggarwal and Sushant Sunder. Writing — original draft: Vidushi Aggarwal and Sushant Sunder. Writing — review & editing: Vidushi Aggarwal, Sushant Sunder and Smita Rastogi Verma. Supervision: Smita Rastogi Verma.

Declarations

Ethics approval Not applicable.

Consent to participate Not applicable.

Consent for publication Not applicable.

Conflict of interest The authors declare no competing interests.

References

Allin KH, Tremaroli V, Caesar R et al (2018) Aberrant intestinal microbiota in individuals with prediabetes. Diabetologia 61:810–820. https://doi.org/10.1007/s00125-018-4550-1

Ambort D, Johansson MEV, Gustafsson JK et al (2012) Calcium and pH-dependent packing and release of the gel-forming MUC2 mucin. Proc Natl Acad Sci 109:5645–5650. https://doi.org/10.1073/pnas.1120269109

Andersson KE, Axling U, Xu J et al (2012) Diverse effects of oats on cholesterol metabolism in C57BL/6 mice correlate with expression of hepatic bile acid-producing enzymes. Eur J Nutr 52:1755–1769. https://doi.org/10.1007/s00394-012-0479-1

Bajer L, Kverka M, Kostovcik M et al (2017) Distinct gut microbiota profiles in patients with primary sclerosing cholangitis and ulcerative colitis. World J Gastroenterol 23:4548. https://doi.org/10.3748/wjg.v23.i25.4548

Bazett M, Bergeron ME, Haston CK (2016) Streptomycin treatment alters the intestinal microbiome, pulmonary T cell profile and airway hyperresponsiveness in a cystic fibrosis mouse model. Sci Rep 6:1–13. https://doi.org/10.1038/srep19189

Bedarf JR, Hildebrand F, Coelho LP et al (2017) Erratum to: functional implications of microbial and viral gut metagenome changes in early stage L-DOPA-naïve Parkinson's disease patients. Genome Med 9:61. https://doi.org/10.1186/s13073-017-0451-z

Belzer C, de Vos WM (2012) Microbes inside—from diversity to function: the case of *Akkermansia*. ISME J 6:1449–1458. https://doi.org/10.1038/ismej.2012.6

Bernardo D, Sánchez B, Al-Hassi HO et al (2012) Microbiota/host crosstalk biomarkers: regulatory response of human intestinal dendritic cells exposed to *Lactobacillus* extracellular encrypted peptide. PLoS ONE 7:e36262. https://doi.org/10.1371/journal.pone.0036262

Caesar R, Tremaroli V, Kovatcheva-Datchary P et al (2015) Crosstalk between gut microbiota and dietary lipids aggravates WAT inflammation through TLR signaling. Cell Metab 22:658–668. https://doi.org/10.1016/j.cmet.2015.07.026

Candela M, Rampelli S, Turroni S et al (2012) Unbalance of intestinal microbiota in atopic children. BMC Microbiol 12:95. https://doi.org/10.1186/1471-2180-12-95

Cani PD, de Vos WM (2017) Next-generation beneficial microbes: the case of *Akkermansia muciniphila*. Front Microbiol. https://doi.org/10.3389/fmicb.2017.01765

Cantarel BL, Waubant E, Chehoud C et al (2015) Gut microbiota in multiple sclerosis. J Investig Med 63:729–734. https://doi.org/10.1097/JIM.0000000000000192

Chelakkot C, Choi Y, Kim D-K et al (2018) *Akkermansia muciniphila*-derived extracellular vesicles influence gut permeability through the regulation of tight junctions. Exp Mol Med 50:e450–e450. https://doi.org/10.1038/emm.2017.282

Chen X, Song S, Zhao Y et al (2020a) 146P A new biomarker of microbe: *Akkermansia muciniphila* in lung adenocarcinoma tissues may predict lymph node metastasis in lung adenocarcinoma. Ann Oncol 31:S297–S298. https://doi.org/10.1016/j.annonc.2020.08.267

Chen Z, Qian X, Chen S et al (2020b) *Akkermansia muciniphila* enhances the antitumor effect of cisplatin in lewis lung cancer mice. J Immunol Res 2020:1–13. https://doi.org/10.1155/2020/2969287

Collado MC, Isolauri E, Laitinen K, Salminen S (2008) Distinct composition of gut microbiota during pregnancy in overweight and normal-weight women. Am J Clin Nutr 88:894–899. https://doi.org/10.1093/ajcn/88.4.894

Collado MC, Laitinen K, Salminen S, Isolauri E (2012) Maternal weight and excessive weight gain during pregnancy modify the immunomodulatory potential of breast milk. Pediatr Res 72:77–85. https://doi.org/10.1038/pr.2012.42

Cone RA (2009) Barrier properties of mucus. Adv Drug Deliv Rev 61:75–85

Cronin P, Joyce SA, O'Toole PW, O'Connor EM (2021) Dietary fibre modulates the gut microbiota. Nutrients 13:1655. https://doi.org/10.3390/nu13051655

Dao MC, Everard A, Aron-Wisnewsky J et al (2016) *Akkermansia muciniphila* and improved metabolic health during a dietary intervention in obesity: relationship with gut microbiome richness and ecology. Gut 65:426–436. https://doi.org/10.1136/gutjnl-2014-308778

De Angelis M, Piccolo M, Vannini L et al (2013) Fecal microbiota and metabolome of children with autism and pervasive developmental disorder not otherwise specified. PLoS ONE 8:e76993. https://doi.org/10.1371/journal.pone.0076993

Dean M, Annilo T (2005) Evolution of the atp-binding cassette (ABC) transporter superfamily in vertebrates. Annu Rev Genomics Hum Genet 6:123–142. https://doi.org/10.1146/annurev.genom.6.080604.162122

Demirci M, Tokman HB, Uysal HK et al (2019) Reduced *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* levels in the gut microbiota of children with allergic asthma. Allergol Immunopathol (Madr) 47:365–371. https://doi.org/10.1016/j.aller.2018.12.009

Depommier C, Everard A, Druart C et al (2019) Supplementation with *Akkermansia muciniphila* in overweight and obese human volunteers: a proof-of-concept exploratory study. Nat Med 25:1096–1103. https://doi.org/10.1038/s41591-019-0495-2

Derrien M, Van Baarlen P, Hooiveld G et al (2011) Modulation of mucosal immune response, tolerance, and proliferation in mice colonized by the mucin-degrader *Akkermansia muciniphila*. Front Microbiol. https://doi.org/10.3389/fmicb.2011.00166

Derrien M, Vaughan EE, Plugge CM, de Vos WM (2004) *Akkermansia municiphila* gen. nov., sp. nov., a human intestinal mucin-degrading bacterium. Int J Syst Evol Microbiol 54:1469–1476. https://doi.org/10.1099/ijs.0.02873-0

Dhar D, Mohanty A (2020) Gut microbiota and COVID-19- possible link and implications. Virus Res 285:198018. https://doi.org/10.1016/j.virusres.2020.198018

Drell T, Larionova A, Voor T et al (2015) Differences in gut microbiota between atopic and healthy children. Curr Microbiol 71:177–183. https://doi.org/10.1007/s00284-015-0815-9

Everard A, Belzer C, Geurts L et al (2013) Cross-talk between *Akkermansia muciniphila* and intestinal epithelium controls diet-induced obesity. Proc Natl Acad Sci 110:9066–9071. https://doi.org/10.1073/pnas.1219451110

Fassatoui M, Lopez-Siles M, Díaz-Rizzolo DA et al (2019) Gut microbiota imbalances in Tunisian participants with type 1 and type 2 diabetes mellitus. Biosci Rep. https://doi.org/10.1042/BSR20182348

Fischer H, Holst E, Karlsson F et al (2015) Altered microbiota in microscopic colitis. Gut 64:1185–1186. https://doi.org/10.1136/gutjnl-2014-308956

Fuke N, Nagata N, Suganuma H, Ota T (2019) Regulation of gut microbiota and metabolic endotoxemia with dietary factors. Nutrients 11:2277. https://doi.org/10.3390/nu11102277

Gao R, Zhu C, Li H et al (2018) Dysbiosis signatures of gut microbiota along the sequence from healthy, young patients to those with overweight and obesity. Obesity 26:351–361. https://doi.org/10.1002/oby.22088

Gargari G, Deon V, Taverniti V et al (2018) Evidence of dysbiosis in the intestinal microbial ecosystem of children and adolescents with primary hyperlipidemia and the potential role of regular hazelnut intake. FEMS Microbiol Ecol. https://doi.org/10.1093/femsec/fiy045

Gou W, Fu Y, Yue L et al (2021) Gut microbiota, inflammation, and molecular signatures of host response to infection. J Genet Genomics 48:792–802. https://doi.org/10.1016/j.jgg.2021.04.002

Grander C, Adolph TE, Wieser V et al (2018) Recovery of ethanol-induced *Akkermansia muciniphila* depletion ameliorates alcoholic liver disease. Gut 67:891–901. https://doi.org/10.1136/gutjnl-2016-313432

Gray J, Oehrle K, Worthen G et al (2017) Intestinal commensal bacteria mediate lung mucosal immunity and promote resistance of newborn mice to infection. Sci Transl Med 9:eaaf9412. https://doi.org/10.1126/scitranslmed.aaf9412

Greer RL, Dong X, Moraes ACF et al (2016) *Akkermansia muciniphila* mediates negative effects of IFNγ on glucose metabolism. Nat Commun 7:13329. https://doi.org/10.1038/ncomms13329

Gunderson KL, Kopito RR (1994) Effects of pyrophosphate and nucleotide analogs suggest a role for ATP hydrolysis in cystic fibrosis transmembrane regulator channel gating. J Biol Chem 269:19349–19353. https://doi.org/10.1016/s0021-9258(17)32174-9

Guo X, Li S, Zhang J et al (2017) Genome sequencing of 39 *Akkermansia muciniphila* isolates reveals its population structure, genomic and functional diversity, and global distribution in mammalian gut microbiotas. BMC Genomics. https://doi.org/10.1186/s12864-017-4195-3

Gurung M, Li Z, You H et al (2020) Role of gut microbiota in type 2 diabetes pathophysiology. EBioMedicine 51:102590. https://doi.org/10.1016/j.ebiom.2019.11.051

Han H, Ma Q, Li C et al (2020) Profiling serum cytokines in COVID-19 patients reveals IL-6 and IL-10 are disease severity predictors. Emerg Microbes Infect 9:1123–1130. https://doi.org/10.1080/22221751.2020.1770129

Hayden H, Eng A, Pope C et al (2019) P314 Fecal dysbiosis is associated with growth failure in infants with cystic fibrosis: a multicentre study. J Cyst Fibros 18:S146. https://doi.org/10.1016/S1569-1993(19)30607-1

Hoffman LR, Pope CE, Hayden HS et al (2014) *Escherichia coli* dysbiosis correlates with gastrointestinal dysfunction in children with cystic fibrosis. Clin Infect Dis 58:396–399. https://doi.org/10.1093/cid/cit715

Hu Y, Feng Y, Wu J et al (2019) The gut microbiome signatures discriminate healthy from pulmonary tuberculosis patients. Front Cell Infect Microbiol. https://doi.org/10.3389/fcimb.2019.00090

Hussain I, Cher GLY, Abid MA, Abid MB (2021) Role of gut microbiome in COVID-19: an insight into pathogenesis and therapeutic potential. Front Immunol. https://doi.org/10.3389/fimmu.2021.765965

Jakobsson HE, Rodríguez-Piñeiro AM, Schütte A et al (2015) The composition of the gut microbiota shapes the colon mucus barrier. EMBO Rep 16:164–177. https://doi.org/10.15252/embr.201439263

Johansson MEV, Holmén Larsson JM, Hansson GC (2011) The two mucus layers of colon are organized by the MUC2 mucin, whereas the outer layer is a legislator of host-microbial interactions. Proc Natl Acad Sci U S A 108:4659–4665. https://doi.org/10.1073/pnas.1006451107

Johansson MEV, Phillipson M, Petersson J et al (2008) The inner of the two Muc2 mucin-dependent mucus layers in colon is devoid of bacteria. Proc Natl Acad Sci U S A 105:15064–15069. https://doi.org/10.1073/pnas.0803124105

Johansson MEV, Sjövall H, Hansson GC (2013) The gastrointestinal mucus system in health and disease. Nat Rev Gastroenterol Hepatol 10:352–361

Karlsson CLJ, Önnerfält J, Xu J et al (2012) The microbiota of the gut in preschool children with normal and excessive body weight. Obesity 20:2257–2261. https://doi.org/10.1038/oby.2012.110

Kim S, Rigatto K, Gazzana MB et al (2020) Altered gut microbiome profile in patients with pulmonary arterial hypertension. Hypertension 75:1063–1071. https://doi.org/10.1161/HYPERTENSIONAHA.119.14294

Kuba K, Imai Y, Rao S et al (2005) A crucial role of angiotensin converting enzyme 2 (ACE2) in SARS coronavirus–induced lung injury. Nat Med 11:875–879. https://doi.org/10.1038/nm1267

Lee DM, Battson ML, Jarrell DK et al (2018) SGLT2 inhibition via dapagliflozin improves generalized vascular dysfunction and alters the gut microbiota in type 2 diabetic mice. Cardiovasc Diabetol 17:62. https://doi.org/10.1186/s12933-018-0708-x

Ley RE, Bäckhed F, Turnbaugh P et al (2005) Obesity alters gut microbial ecology. Proc Natl Acad Sci 102:11070–11075. https://doi.org/10.1073/pnas.0504978102

Li J, Lin S, Vanhoutte PM et al (2016) *Akkermansia muciniphila* protects against atherosclerosis by preventing metabolic endotoxemia-induced inflammation in apoe −/− mice. Circulation 133:2434–2446. https://doi.org/10.1161/CIRCULATIONAHA.115.019645

Liu F, Ling Z, Xiao Y et al (2017) Dysbiosis of urinary microbiota is positively correlated with Type 2 diabetes mellitus. Oncotarget 8:3798–3810. https://doi.org/10.18632/oncotarget.14028

Lukovac S, Belzer C, Pellis L et al (2014) Differential modulation by *Akkermansia muciniphila* and *Faecalibacterium prausnitzii* of host peripheral lipid metabolism and histone acetylation in mouse gut organoids. Mbio. https://doi.org/10.1128/mBio.01438-14

Malham M, Lilje B, Houen G et al (2019) The microbiome reflects diagnosis and predicts disease severity in paediatric onset inflammatory bowel disease. Scand J Gastroenterol 54:969–975. https://doi.org/10.1080/00365521.2019.1644368

Manor O, Levy R, Pope CE et al (2016) Metagenomic evidence for taxonomic dysbiosis and functional imbalance in the gastrointestinal tracts of children with cystic fibrosis. Sci Rep 6:22493. https://doi.org/10.1038/srep22493

Marvasti FE, Moshiri A, Sadat Taghavi M et al (2020) The first report of differences in gut microbiota composition between obese and normal weight iranian subjects. Iran Biomed J 24:148–154. https://doi.org/10.29252/ibj.24.3.148

Mattioli AV, Sciomer S, Cocchi C et al (2020) Quarantine during COVID-19 outbreak: changes in diet and physical activity increase the risk of cardiovascular disease. Nutr Metab Cardiovasc Dis 30:1409–1417. https://doi.org/10.1016/j.numecd.2020.05.020

Meeker SM, Mears KS, Sangwan N et al (2020) CFTR dysregulation drives active selection of the gut microbiome. PLoS Pathog 16:e1008251. https://doi.org/10.1371/journal.ppat.1008251

Michalovich D, Rodriguez-Perez N, Smolinska S et al (2019) Obesity and disease severity magnify disturbed microbiome-immune interactions in asthma patients. Nat Commun 10:5711. https://doi.org/10.1038/s41467-019-13751-9

Mira-Pascual L, Cabrera-Rubio R, Ocon S et al (2015) Microbial mucosal colonic shifts associated with the development of colorectal cancer reveal the presence of different bacterial and archaeal biomarkers. J Gastroenterol 50:167–179. https://doi.org/10.1007/s00535-014-0963-x

Morgan XC, Huttenhower C (2012) Chapter 12: human microbiome analysis. PLoS Comput Biol 8:e1002808. https://doi.org/10.1371/journal.pcbi.1002808

Neef A, Sanz Y (2013) Future for probiotic science in functional food and dietary supplement development. Curr Opin Clin Nutr Metab Care 16:679–687. https://doi.org/10.1097/MCO.0b013e328365c258

Nicholson JK, Holmes E, Kinross J et al (2012) Host-gut microbiota metabolic interactions. Science 336:1262–1267. https://doi.org/10.1126/science.1223813

Opstelten JL, Plassais J, van Mil SWC et al (2016) Gut microbial diversity is reduced in smokers with Crohn's disease. Inflamm Bowel Dis 22:2070–2077. https://doi.org/10.1097/MIB.0000000000000875

Ottman N, Davids M, Suarez-Diez M et al (2017a) Genomescale model and omics analysis of metabolic capacities of *Akkermansia muciniphila* reveal a preferential mucin-degrading lifestyle. Appl Environ Microbiol. https://doi.org/10.1128/AEM.01014-17

Ottman N, Reunanen J, Meijerink M et al (2017b) Pili-like proteins of *Akkermansia muciniphila* modulate host immune responses and gut barrier function. PLoS ONE. https://doi.org/10.1371/journal.pone.0173004

Ouwerkerk JP, van der Ark KCH, Davids M et al (2016) Adaptation of *Akkermansia muciniphila* to the oxic-anoxic interface of the mucus layer. Appl Environ Microbiol 82:6983–6993. https://doi.org/10.1128/AEM.01641-16

Pang T, Leach ST, Katz T et al (2014) Fecal biomarkers of intestinal health and disease in children. Front Pediatr. https://doi.org/10.3389/fped.2014.00006

Peng Y, Zhang S, Liu Z et al (2020) Gut microbiota and Chinese medicine syndrome: altered fecal microbiotas in spleen (Pi)-deficient patients. J Tradit Chin Med = Chung i Tsa Chih Ying Wen Pan 40:137–143

Plovier H, Everard A, Druart C et al (2016) A purified membrane protein from *Akkermansia muciniphila* or the pasteurized bacterium improves metabolism in obese and diabetic mice. Nat Med 23:107–113. https://doi.org/10.1038/nm.4236

Png CW, Lindén SK, Gilshenan KS et al (2010) Mucolytic bacteria with increased prevalence in IBD mucosa augment in vitro utilization of mucin by other bacteria. Am J Gastroenterol 105:2420–2428. https://doi.org/10.1038/ajg.2010.281

Price CE, O'Toole GA (2021) The gut-lung axis in cystic fibrosis. J Bacteriol 203

Qin J, Li Y, Cai Z et al (2012) A metagenome-wide association study of gut microbiota in type 2 diabetes. Nature 490:55–60. https://doi.org/10.1038/nature11450

Rajilić-Stojanović M, Shanahan F, Guarner F, de Vos WM (2013) Phylogenetic analysis of dysbiosis in ulcerative colitis during remission. Inflamm Bowel Dis 19:481–488. https://doi.org/10.1097/MIB.0b013e31827fec6d

Remely M, Hippe B, Geretschlaeger I et al (2015) Increased gut microbiota diversity and abundance of *Faecalibacterium prausnitzii* and *Akkermansia* after fasting: a pilot study. Wien Klin Wochenschr 127:394–398. https://doi.org/10.1007/s00508-015-0755-1

Remely M, Hippe B, Zanner J et al (2016) Gut microbiota of obese, type 2 diabetic individuals is enriched in *Faecalibacterium prausnitzii*, *Akkermansia muciniphila* and *Peptostreptococcus anaerobius* after weight loss. Endocrine, Metab Immune Disord - Drug Targets 16:99–106. https://doi.org/10.2174/1871530316666160831093813

Ruiz L, Hevia A, Bernardo D et al (2014) Extracellular molecular effectors mediating probiotic attributes. FEMS Microbiol Lett 359:1–11. https://doi.org/10.1111/1574-6968.12576

Salonen A, de Vos WM (2014) Impact of diet on human intestinal microbiota and health. Annu Rev Food Sci Technol 5:239–262. https://doi.org/10.1146/annurev-food-030212-182554

Sánchez B, Bressollier P, Urdaci MC (2008) Exported proteins in probiotic bacteria: adhesion to intestinal surfaces, host immunomodulation and molecular cross-talking with the host. FEMS Immunol Med Microbiol 54:1–17. https://doi.org/10.1111/j.1574-695x.2008.00454.x

Sánchez B, Urdaci MC, Margolles A (2010) Extracellular proteins secreted by probiotic bacteria as mediators of effects that promote mucosa–bacteria interactions. Microbiology 156:3232–3242. https://doi.org/10.1099/mic.0.044057-0

Schippa S, Iebba V, Santangelo F et al (2013) Cystic fibrosis transmembrane conductance regulator (CFTR) allelic variants relate to shifts in faecal microbiota of cystic fibrosis patients. PLoS ONE 8:e61176. https://doi.org/10.1371/journal.pone.0061176

Schneeberger M, Everard A, Gómez-Valadés AG et al (2015) *Akkermansia muciniphila* inversely correlates with the onset of inflammation, altered adipose tissue metabolism and metabolic disorders during obesity in mice. Sci Rep 5:16643. https://doi.org/10.1038/srep16643

Schütte A, Ermund A, Becker-Pauly C et al (2014) Microbial-induced meprin β cleavage in MUC2 mucin and a functional CFTR channel are required to release anchored small intestinal mucus. Proc Natl Acad Sci 111:12396–12401. https://doi.org/10.1073/pnas.1407597111

Sedighi M, Razavi S, Navab-Moghadam F et al (2017) Comparison of gut microbiota in adult patients with type 2 diabetes and healthy individuals. Microb Pathog 111:362–369. https://doi.org/10.1016/j.micpath.2017.08.038

Shin J, Noh JR, Chang DH et al (2019) Elucidation of *Akkermansia muciniphila* probiotic traits driven by mucin depletion. Front Microbiol. https://doi.org/10.3389/fmicb.2019.01137

Shin N-R, Lee J-C, Lee H-Y et al (2014) An increase in the *Akkermansia spp.* population induced by metformin treatment improves glucose homeostasis in diet-induced obese mice. Gut 63:727–735. https://doi.org/10.1136/gutjnl-2012-303839

Snider EJ, Compres G, Freedberg DE et al (2019) Alterations to the esophageal microbiome associated with progression from Barrett's esophagus to esophageal adenocarcinoma. Cancer Epidemiol Biomarkers Prev 28:1687–1693. https://doi.org/10.1158/1055-9965.EPI-19-0008

Swidsinski A, Dörffel Y, Loening-Baucke V et al (2011) Acute appendicitis is characterised by local invasion with *Fusobacterium nucleatum/necrophorum*. Gut 60:34–40. https://doi.org/10.1136/gut.2009.191320

Tan L, Zhao S, Zhu W et al (2018) The *Akkermansia muciniphila* is a gut microbiota signature in psoriasis. Exp Dermatol 27:144–149. https://doi.org/10.1111/exd.13463

Thomsson KA, Hinojosa-Kurtzberg M, Axelsson KA et al (2002) Intestinal mucins from cystic fibrosis mice show increased fucosylation due to an induced Fucα1-2 glycosyltransferase. Biochem J 367:609–616. https://doi.org/10.1042/BJ20020371

Tilg H, Moschen AR (2014) Microbiota and diabetes: an evolving relationship. Gut 63:1513–1521. https://doi.org/10.1136/gutjnl-2014-306928

Tummler B, Puchelle E (1997) CFTR: a multifaceted epithelial molecule. Trends Cell Biol 6(7):250–251

Unger MM, Spiegel J, Dillmann K-U et al (2016) Short chain fatty acids and gut microbiota differ between patients with Parkinson's

disease and age-matched controls. Parkinsonism Relat Disord 32:66–72. https://doi.org/10.1016/j.parkreldis.2016.08.019

Vakili B, Fateh A, Asadzadeh Aghdaei H et al (2020) Characterization of gut microbiota in hospitalized patients with clostridioides difficile infection. Curr Microbiol 77:1673–1680. https://doi.org/10.1007/s00284-020-01980-x

van der Ark KCH, Aalvink S, Suarez-Diez M et al (2018) Model-driven design of a minimal medium for *Akkermansia muciniphila* confirms mucus adaptation. Microb Biotechnol 11:476–485. https://doi.org/10.1111/1751-7915.13033

van Passel MWJ, Kant R, Zoetendal EG et al (2011) The genome of *Akkermansia muciniphila*, a dedicated intestinal mucin degrader, and its use in exploring intestinal metagenomes. PLoS ONE. https://doi.org/10.1371/journal.pone.0016876

Vernocchi P, Del CF, Russo A et al (2018) Gut microbiota signatures in cystic fibrosis: Loss of host CFTR function drives the microbiota enterophenotype. PLoS ONE 13:e0208171. https://doi.org/10.1371/journal.pone.0208171

Wan L, Zhou X, Wang C et al (2019) Alterations of the gut microbiota in multiple system atrophy patients. Front Neurosci. https://doi.org/10.3389/fnins.2019.01102

Wang H, Wang H, Sun Y et al (2021) Potential associations between microbiome and COVID-19. Front Med. https://doi.org/10.3389/fmed.2021.785496

Wang J, Jia H (2016) Metagenomewide association studies: fine-mining the microbiome. Nat Rev Microbiol 14:508–522. https://doi.org/10.1038/nrmicro.2016.83

Wang J, Tang H, Zhang C et al (2015) Modulation of gut microbiota during probiotic-mediated attenuation of metabolic syndrome in high fat diet-fed mice. ISME J 9:1–15. https://doi.org/10.1038/ismej.2014.99

Wang L, Christophersen CT, Sorich MJ et al (2011) Low relative abundances of the mucolytic bacterium *Akkermansia muciniphila* and *Bifidobacterium spp.* in feces of children with autism. Appl Environ Microbiol 77:6718–6721. https://doi.org/10.1128/AEM.05212-11

Wang Z, Saha S, Van Horn S et al (2018) Gut microbiome differences between metformin- and liraglutide-treated T2DM subjects. Endocrinol Diabetes Metab 1:e00009. https://doi.org/10.1002/edm2.9

Weir TL, Manter DK, Sheflin AM et al (2013) Stool microbiome and metabolome differences between colorectal cancer patients and healthy adults. PLoS ONE 8:e70803. https://doi.org/10.1371/journal.pone.0070803

Welsh MJ, Smith AE (1993) Molecular mechanisms of CFTR chloride channel dysfunction in cystic fibrosis. Cell 73:1251–1254

Xiao F, Sun J, Xu Y et al (2020) Infectious SARS-CoV-2 in feces of patient with severe COVID-19. Emerg Infect Dis 26:1920–1922. https://doi.org/10.3201/eid2608.200681

Xu R, Wu B, Liang J et al (2020a) Altered gut microbiota and mucosal immunity in patients with schizophrenia. Brain Behav Immun 85:120–127. https://doi.org/10.1016/j.bbi.2019.06.039

Xu Y, Wang N, Tan H-Y et al (2020b) Function of *Akkermansia muciniphila* in obesity: interactions with lipid metabolism, immune response and gut systems. Front Microbiol. https://doi.org/10.3389/fmicb.2020.00219

Yamamoto S, Saito M, Tamura A, Prawisuda D, Mizutani T, Yotsuyanagi H (2021) The human microbiome and COVID-19: a systematic review. PLoS ONE 16(6):e0253293. https://doi.org/10.1371/journal.pone.0253293

Yeoh YK, Zuo T, Lui GC-Y et al (2021) Gut microbiota composition reflects disease severity and dysfunctional immune responses in patients with COVID-19. Gut 70:698–706. https://doi.org/10.1136/gutjnl-2020-323020

Yu J, Feng Q, Wong SH et al (2017) Metagenomic analysis of faecal microbiome as a tool towards targeted non-invasive biomarkers for colorectal cancer. Gut 66:70–78. https://doi.org/10.1136/gutjnl-2015-309800

Zhang L, Qin Q, Liu M et al (2018) *Akkermansia muciniphila* can reduce the damage of gluco/lipotoxicity, oxidative stress and inflammation, and normalize intestine microbiota in streptozotocin-induced diabetic rats. Pathog Dis. https://doi.org/10.1093/femspd/fty028

Zhong H, Ren H, Lu Y et al (2019) Distinct gut metagenomics and metaproteomics signatures in prediabetics and treatment-naïve type 2 diabetics. EBioMedicine 47:373–383. https://doi.org/10.1016/j.ebiom.2019.08.048

Zhu F, Ju Y, Wang W et al (2020) Metagenome-wide association of gut microbiome features for schizophrenia. Nat Commun 11:1612. https://doi.org/10.1038/s41467-020-15457-9

Zoetendal EG, Raes J, van den Bogert B et al (2012) The human small intestinal microbiota is driven by rapid uptake and conversion of simple carbohydrates. ISME J 6:1415–1426. https://doi.org/10.1038/ismej.2011.212

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

International Journal of Multidisciplinary Research and Growth Evaluation.

Earth-air heat exchanger and its utilization in cold storages

Palak Kumari [1*], Martin Leo King [2], Dr. JP Kesari [3]
[1-2] Student, Delhi Technological University, Formerly Delhi College of Engineering, Bawana Road, Delhi, India
[3] Associate Professor, Delhi Technological University, Formerly Delhi College of Engineering, Bawana Road, Delhi, India

* Corresponding Author: **Palak Kumari**

Article Info

ISSN (online): 2582-7138
Volume: 03
Issue: 02
March-April 2022
Received: 22-03-2022;
Accepted: 12-04-2022
Page No: 564-569

Abstract
Cold Storage is a special type of room, the temperature is kept very low with the help of machinery and precision equipment. Marine products are also produced in large quantities due to the large coastal areas. Therefore, it is a good job for entrepreneurs to invest. Uses Cold space is used to save fruit and vegetables. If kept in a cool place, they will not rot after many months. Sometimes, during the production of certain vegetable or fruit crops, the demand for those items decreases, which in turn reduces the consumption of the residual item and is stored in a cool place. Cold storage works primarily on the 'vapor pressure cycle' to produce cooling. Now, we want to introduce a better, more efficient, and less expensive method EAHE (Earth-Warm Temperature). The Earth-air heat exchanger (EAHE) is a promising method that can be used effectively to reduce the heat / cooling load of a building by first cooling the air in summer. Over the past two decades, much research has been done to develop analytical models and analysis numbers for EAHE systems. The method of calculating the fixed temperature of the earth (EUT) and the recently developed correlation of friction factor and Nusselt number are used to ensure maximum accuracy in calculating heat transfer. Advanced calculations allow designers to calculate heat transfer, coefficient of flexible heat transfer, pressure drop, and length of the EAHE system pipeline. A long pipe of small diameter buried at high depth and having a low flow rate of air causes an increase in the efficiency of the EAHE system.

Keywords: cold storages, exchanger, Earth-air, temperature

Introduction
Over the past two decades, much research has been done to develop analytical and numerical models to analyze EAHE systems (Mihalakakou *et al*. 1994; Bojic *et al*. 1997; Gauthier *et al*. 1997; Hollmuller and Lachal 2001; Su *et al*. 2012; Sehli *et al*. 2012; Ozgener *et al*. 2013) [9]. The performance analysis of EAHE involved the calculation of the continuous heat transfer from the pipe to the substrate or the calculation of the variable heat transfer from the circulating air to the pipe and the changes in air temperature and humidity. A number of computer modeling tools are commercially available. Energy Plus and TRSYS have well-functioning EAHE modules; however, these are analytical tools and are not used for immediate design.
Currently, Computational Fluid Dynamics (CFD) is best known to researchers for modeling and evaluating the performance of EAHE systems. CFD uses a very simple rule to divide the entire system into smaller grids. Then, governing standards were applied to these different factors to obtain numerical solutions for motion parameters, pressure distribution, and temperature gradients in the short term and at reasonable cost due to the reduction of the required test function (Kanaris *et al*. 2006; Wang *et al*. 2007). A thorough analysis of the EAHE system, the use of CFD is recommended, but limited to those with a good command over it. For the first design of the EAHE system, the use of basic heat transfer calculators is best suited to determine the geometric size of the system.

Many researchers such as De Paepe and Janssens (2003), Badescu and Isvoranu (2011) [2], and T'Joen *et al*. (2012) developed EAHE design equations and processes.
In this paper, the author has developed a one-sided model of the EAHE system. The EUT calculation method and the recently developed correlation of friction factor and Nusselt number are used to ensure maximum accuracy in calculating heat transfer.

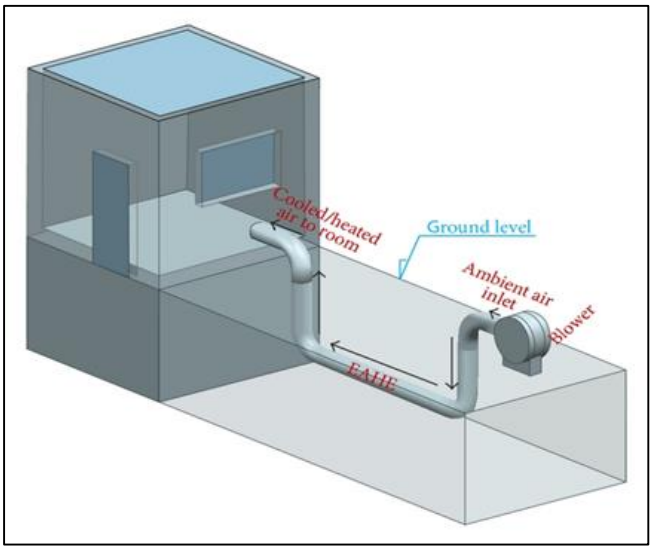


Fig 1

Review of technologies
The development of the EAHE system model involves the use of basic heat transfer calculators. The geometric measurements of the EAHE system are determined by considering the amount of heat or cooling that will be met to adjust the building surface. The design process involves identifying user input parameters as well as parameters that affect the output of the desired design. Once the output output has been adjusted, heat transfer gauges are used to meet the output output in terms of input parameters.
The geometric sizing parameters of an EAHE comprises the diameter of the pipe, D; length of the pipe, L; and number of pipes in parallel, N_p, in the heat exchanger.

Assumptions
The following assumptions are made to facilitate the development of a one-sided EAHE system model:

- The surface temperature is defined as the relative temperature of the local air, which is equal to the temperature of the air intake.
- EUT can be measured at the average annual temperature of the region (Bhopal, India).
- The polyvinyl chloride (PVC) pipe used in EAHE is a uniform cross-section.
- The size of the pipeline used in EAHE is very small; therefore, the heat resistance of the pipe material is not taken into account.
- The temperature on the surface of the pipe is the same as the axial direction.

Boundary conditions
The following boundary conditions were used in the one-dimensional model of the EAHE system.
Inlet boundary conditions
At the entrance to the EAHE pipe, the values of air flow velocity, v_a (m/s), and static temperature of air, T_{in} (°C), at inlet were to be defined. The thermodynamic properties (density and specific heat capacity) and transport properties (dynamic viscosity and thermal conductivity) of air were to be defined at static temperature of air at inlet.

Outlet boundary conditions
In a subsonic flow regime, the relative pressure at the outlet of the EAHE pipe was defined as equal to zero atm.

Wall
The temperature on the surface of pipe (wall) was uniform in axial direction and was defined as equal to earth's undisturbed temperature at Bhopal city (25.2 °C). No slip condition with smooth wall was assumed at the inner surface of the pipe.

Mass flow rate of air
The mass air flow rate is an important factor, and should be known by the designer in order to determine the choice of size and number of pipes. No unique size and pipeline size can meet EAHE performance. Therefore, the designer should consider the best combination of EAHE performance and pumping power required to ensure mass air flow. For a pipe of diameter, D; air density, ρ; air flow velocity, v_a; and number of parallel pipes, N_p, the mass flow rate of air through a pipe, m_a, is given by:

$$\dot{m} = \frac{\frac{\pi}{4} D^2 \rho v_a}{N_p} \quad (1)$$

For the designer, these parameters have to be determined in such a way that the boundary conditions and the heat exchanger performance are met.
Earth's undisturbed temperature
The uninterrupted global temperature is an important parameter in designing an EAHE system. Assuming that the homogeneous soil of continuous thermal diffusivity, the temperature at any depth z and time t can be measured by the following expression (Labs 1989):

$$T_{z,t} = T_m - A_s exp\left[-z\left(\frac{\pi}{365\alpha_s}\right)^{\frac{1}{2}}\right] \cos\left\{\frac{2\pi}{365}\left[t - t_o - \frac{z}{2}\left(\frac{365}{\pi\alpha_s}\right)^{\frac{1}{2}}\right]\right\} \quad (2)$$

where $T_{z,t}$ is representing the ground temperature at time t (s) and depth z (m), T_m is the average soil surface temperature (°C), A_s is the amplitude of soil surface variation (°C), α_s is the soil thermal diffusivity (m^2/s; m^2/day), t is the time elapsed from beginning of the calendar year (day), and t_o is the phase constant of soil surface (s; days). It is very difficult to accurately calculate the uninterrupted temperature of the earth because the boundaries of the soil are often unknown. In addition, the central soil structures are described. Therefore, the uninterrupted global warming is an estimated value that can be considered to be equal to the annual average of global warming. The surface temperature is considered to be equal to the prevailing air temperature. Thus, the uninterrupted global temperature of Bhopal (Central India) is defined as 25.2 ° C equal to the average annual temperature (source: Department of Meteorology, Bhopal).

Methods

If the of the EAHE system is known, the calculation of the heat transfer rate can be done using the log mean temperature difference method (LMTD) or the ε – number of transfer units (NTU) method. In this paper the ε – NTU method is used. Exhaust temperature is determined using the efficiency of EAHE (ε) which is the function of the number of transmission units (NTU).

Heat exchanger effectiveness and NTU

In the system of earth–air heat exchanger, the area used to transport air heaters only. Heat is released or pumped air into the pipe walls by moving and from the pipe walls to the surrounding soil and vice versa by the conveyor. If the contact of the pipe wall with the ground is considered complete and the groundwater flow is considered to be very high compared to the high resistance, then the temperature of the wall inside the pipe can be considered permanent. The NTU discourse relies on different types of EAHE system flow configurations. In this paper, an evaporator or condenser (with a constant temperature on one side, i.e., a wall) is used. The total heat transferred to the air when flowing through a buried pipe is given by:

$$Q_{\mathrm{h}} = \dot{m} C_{\mathrm{p}} \left(T_{\mathrm{out}} - T_{\mathrm{in}}\right) \tag{3}$$

where $\dot{m}$ is the mass flow rate of air (kg/s), C_{p} is the specific heat of air (J/kg-K), T_{out} is the temperature of air at outlet of EAHE pipe (°C), and T_{in} is the temperature of air at inlet of EAHE pipe (°C).

Due to convection between the wall and the air, the transferred heat can also be given by:

$$Q_{\mathrm{h}} = hA\Delta T_{\mathrm{lm}} \tag{4}$$

where h is the convective heat transfer coefficient (W/m^2-K) and A is the internal surface area of the pipe (m^2).

The logarithmic average temperature difference (ΔT_{lm}) is given by ($T_{\mathrm{EUT}} = T_{\mathrm{wall}}$):

$$\Delta T_{\mathrm{lm}} = \frac{T_{\mathrm{in}} - T_{\mathrm{out}}}{\ln\left[\frac{(T_{\mathrm{in}} - T_{\mathrm{wall}})}{(T_{\mathrm{out}} - T_{\mathrm{wall}})}\right]} \tag{5}$$

The temperature of air at the outlet of the EAHE pipe can be obtained in an exponential form as a function of the wall temperature and inlet air temperature by eliminating Q_{h} from Eqs. (3) and (4).

$$T_{\mathrm{out}} = T_{\mathrm{wall}} + (T_{\mathrm{in}} - T_{\mathrm{wall}})\, e^{-\left(hA/\dot{m}C_{\mathrm{p}}\right)} \tag{6}$$

If a pipe of infinite length ($A = \infty$) is used, the air will be heated or cooled to the wall temperature. The effectiveness (ε) of EAHE for winter heating application can thus be defined as:

$$\varepsilon = \frac{T_{\mathrm{out}} - T_{\mathrm{in}}}{T_{\mathrm{wall}} - T_{\mathrm{in}}} = 1 - e^{-\left(hA/\dot{m}C_{\mathrm{p}}\right)} \tag{7}$$

The non-dimensional group is called the number of transfer units (NTU):

$$\mathrm{NTU} = \frac{hA}{\dot{m}C_{\mathrm{p}}} \tag{8}$$

Which gives

$$\varepsilon = 1 - e^{-\mathrm{NTU}} \tag{9}$$

The operation of the earth-air heating system is determined by the flawless NTU group. The variance in the efficiency of the earth-air temperature switch as the function of the number of transmission units is shown in Figure 2. It was noted that with the increase in NTU value, efficiency also increases but the curve decreases rapidly. The associated profitability in its operation is very small after the NTU value exceeds 3. There are several ways to create a global air temperature changer to get a given NTU as well as a desirable performance. Similar results have been observed by De Paepe and Janssens (2003).

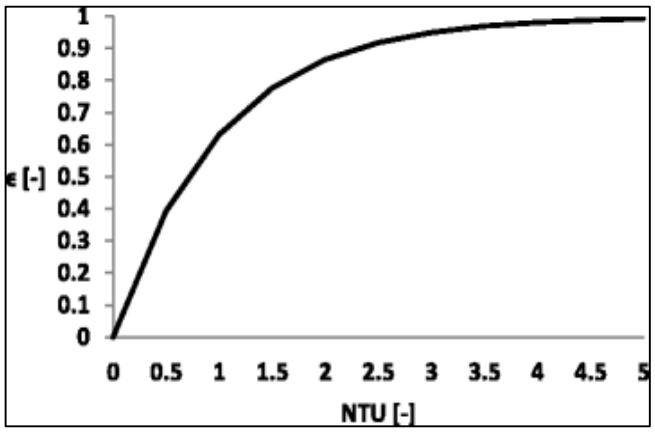


Fig 2: Earth–air heat exchanger effectiveness as a function of number of transfer units

The impact of design parameters on NTU can be studied by heat transfer and pressure reduction. The NTU consists of three parameters, namely, the convective heat transfer coefficient (h), the internal area of the pipe (A) and the air flow rate (ṁ) which can vary.

The inner surface of the pipe is a function of the width, D, and length of the EAHE pipe, L, both:

$$A = \pi D L \tag{10}$$

The convective heat transfer coefficient inside the pipe is defined by:

$$h = \frac{N_{\mathrm{u}} K}{D} \tag{11}$$

where K is the thermal conductivity (W/m-K).

Zhang (2009) presented in his PhD thesis that in conventional earth-to-air heat exchanger (ETAHE) systems, it is typical to have buried ducts with 10 cm $< D_h <$ 40 cm and lengths longer than 20 m. Such sizes mean the ratios of the lengths to the hydraulic diameters (D_h) are at an order of magnitude of 100. The hydraulic diameter is defined as four times the ratio of the cross-section area to wetted perimeter of the cross-section.

$$D_h = \frac{4A}{P} \tag{12}$$

where A is the cross-section area and P is the wetted perimeter of the cross-section

The hydraulic diameter of a circular tube is simply the diameter of the tube. Therefore, it is reasonable to assume that air flow is fully optimized for EAHE of such sizes and to adapt to the corresponding empirical relationship to calculate the convection heat transfer coefficient (CHTC). To test this assumption, eight Nusselt (N u) numbers are related to other ETAHE simulation studies (Arzano and Goswami 1997; Bojic *et al*. 1997, cooling and heating; Singh 1994; De Paepe and Janssens 2003; Hollmuller 2003; Sodha; *et al*. 1994; Benkert and Heidt 1997) [1, 9, 3] were used. Since all correlations were found to fully improve the flow of turbulent air, appropriately, it is expected to produce the same values in the same operating condition. Nusselt's numerical variance with respect to Reynold's standard ETAHE design was drawn using all eight links to calculate CHTC, and the largest difference was observed among the eight correlation outcomes. This may be due to a variety of test conditions, which are adopted to determine the correlation, for example, the overlap of the test channels. Significant differences indicate that the appropriate correlation should be chosen when one uses any available models to simulate the performance of the EAHE system.

The EAHE system described in this paper contains cylindrical pipes of 0.1016 m in diameter made of PVC with a total burial length of 19.228 m. Assuming that the internal surface area of the PVC pipes used in the EAHE system is smooth, the N u links provided by De and Janssens (2003) can be used to measure system performance.

$$N_u = \frac{f/8\,(R_e - 1000)\,P_r}{1 + 12.7\sqrt{(f/8)}\left(P_r^{2/3} - 1\right)} \tag{13}$$

(For turbulent flow in tubes with smooth internal surface)

where R_e is the Reynolds number, P_r is Prandtl number, and f is the friction factor for smooth pipes

With $f = (1.82 \log R_e - 1.64)^{-2}$ (14)

If $2300 \leq R_e < 5 \times 10^6$ and $0.5 < P_r < 10^6$

The Reynolds number is related to the average air velocity and diameter:

$$R_e = \frac{\rho v_a D}{\mu} \tag{15}$$

where v_a is the velocity of air through pipe (m/s), D is the diameter of the pipe (m), and μ is the dynamic viscosity of air (kg/m-s).

The Prandtl number is given by

$$P_r = \frac{\mu c_p}{K} \tag{16}$$

where c_p is the specific heat of air (J/kg-K)

Results and discussion

The thermo-physical properties of materials used in design calculations of EAHE needed to be installed in the cold storages are shown below:

Table 1 Thermo-physical properties of materials used in design calculations of EAHE

The value of Reynolds number was calculated for air flow velocities of 2, 3.5, and 5 m/s and thermo-physical properties of air at 16.7 °C for winter heating application. The corresponding values of friction factor were calculated by using Eq. (14). The value of Prandtl number for thermo-physical properties of air at 16.7 °C was calculated as 0.717. After evaluation of Reynolds number, friction factor, and Prandtl number, the Nusselt number was calculated using Eq. (13) corresponding to air flow velocities of 2, 3.5, and 5 m/s. The variation of Nusselt number with respect to Reynolds number is shown in Fig. 3. It was observed that the value of Nusselt number increases with increase in Reynolds number. Many other researchers (Nakamura and Tamotsu 2004; Luciu *et al*. 2009) also found that the value of Nusselt number increases with increase in Reynolds number.

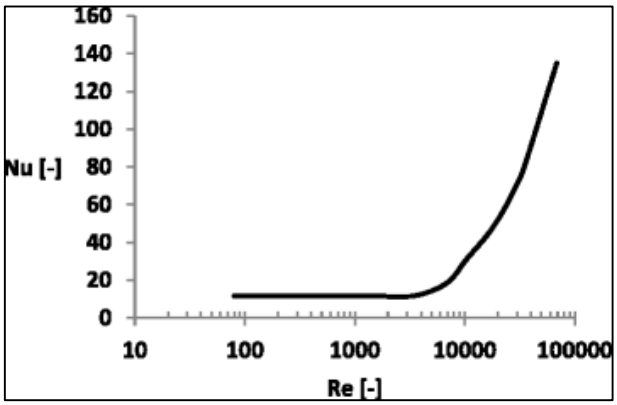


Fig 3: Variation of Nusselt number with respect to Reynolds number

The CHTC for different air flow velocities was calculated by using Eq. (11). The CHTC increases with increase in air flow velocity as shown in Fig. 4. Xiao *et al*. (2011) found that the variation in value of CHTC with a change in air flow velocity has shown similar trends.

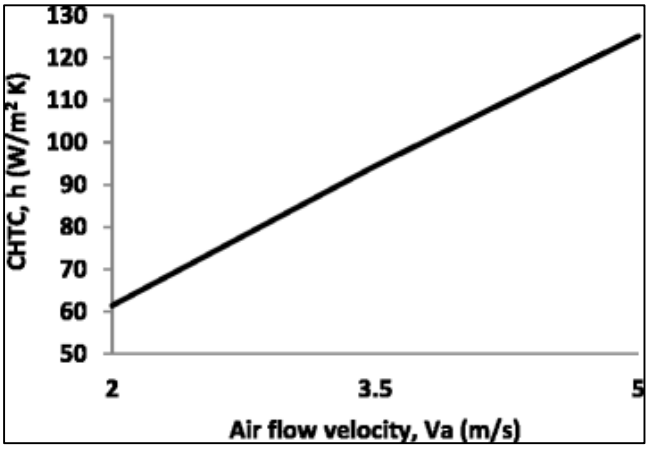


Fig 4: Convective heat transfer coefficient vs. air flow velocity

The length, *L*, is an independent parameter influencing the NTU. There is a linear variation of NTU with length. Changing either the diameter, *D*, or mass flow rate, $\dot{m}$, changes the air velocity inside the tube. This result in a changing Reynolds number, R_e. *D* and $\dot{m}$ have thus no independent influence on the NTU. In general, lowering *D* raises the effectiveness; higher flow rates reduce the effectiveness. So, it is better to have several tubes of small diameter over which the flow rate is divided. Long tubes with a small diameter are profitable for the heat transfer. They, however, raise the pressure drop in the tubes, resulting in high fan energy.

Influence on pressure drop

The pressure drop in a smooth pipe is given by

$$\Delta p = f\frac{L}{D}\rho\frac{{v_a}^2}{2} \tag{17}$$

$$f = \frac{64}{R_e} \text{ If } R_e < 2300$$

$$f = (1.82\log R_e - 1.64)^{-2} \; If \; R_e \geq 2300$$

It is noted from Eqs. (8) and (17) that both NTU and Δ*p* are proportional to the length of the pipe, and the designer can use NTU/*L* and Δ*p*/*L* as the main performance measures to determine the required length of pipe for design purposes. The length of pipe, *L*, is an independent parameter which has a linear influence on pressure drop. The diameter of the pipe and air flow velocity has a combined effect on pressure drop. The decrease in air flow velocity and increase in diameter of pipe results in decrease in pressure drop. This is in disagreement with the thermal demand of a small diameter. In each case, a large number of pipes are beneficial. The combination of pipe length and diameter has to be optimized. To evaluate the overall thermo-hydraulic performance of a specific configuration, the J-factor, introduced by De Paepe and Janssens (2003) was used. It is the ratio of the pressure drop to the NTU value and was shown to be a good performance metric of an earth–air heat exchanger.

$$J = \frac{\Delta P}{\text{NTU}} \tag{18}$$

Conclusions

The earth–air heat exchanger is a promising technique which can effectively be used in the cold storages to provide the cooled air in summer. Many researchers have developed EAHE design equations and procedures. For a complete analysis of the EAHE system, the use of CFD is recommended but it is limited to those who have a good command over it. For the initial design of an EAHE system, the use of basic heat transfer equations is more suitable to determine the geometrical dimensions of the system. In this paper, the author has developed a one-dimensional model of the EAHE system. The method to calculate the EUT and more recently developed correlations for friction factor and Nusselt number are used to ensure higher accuracy in the calculation of heat transfer. The value of EUT for Bhopal (Central India) was calculated as 25.2 °C. It was observed that Nusselt number increases with increase in Reynolds number.

The design of earth–air heat exchanger mainly depends on the heating/cooling load requirement of a building to be conditioned. After calculation of heating/cooling load, the design of the earth–air heat exchanger only depends on the geometrical constraints and cost analysis. The diameter of pipe, pipe length, and number of pipes are the main parameters to be determined. With an increase in length of pipe, both pressure drop and thermal performance increase. A longer pipe of smaller diameter buried at a greater depth and having lower air flow velocity results in an increase in performance of the EAHE system.

References

1. Arzano L, Goswami DY. Performance Analysis of a Closed Loop Underground Air Tunnel for Residential Housing in a Hot and Humid Climate. Proceedings of the ISES Solar World Congress, Taejon, South Korea, August, 1997.
2. Badescu V, Isvoranu D. Pneumatic and thermal design procedure and analysis of earth-to-air heat exchangers of registry type. Appl Energy. 2011; 88:1266-1280.
3. Benkert S, Heidt FD, Scholer D. Calculation tool for Earth heat exchangers GAEA. Proceeding of Building Simulation, Fifth International IBPSA Conference, Prague, Czech Republic, 1997, 9-16.
4. Bisoniya TS, Kumar A, Baredar P. Experimental and analytical studies of earth–air heat exchanger (EAHE) systems in India: a review. Renew Sustain Energy Rev. 2013; 19:238-246.
5. Bisoniya TS, Kumar A, Baredar P. Cooling potential evaluation of earth–air heat exchanger system for summer season. Int J Eng Tech Res. 2014a; 2(4):309-316.
6. Bisoniya TS, Kumar Anil, Baredar Prashant. Heating potential evaluation of earth–air heat exchanger system for winter season. Journal of Building Physics, 2014b, 1-19. Doi: 10.1177/1744259114542403.
7. Bisoniya TS, Kumar A, Baredar P. Parametric analysis of Earth-air heat exchanger system based on CFD

modelling. Int J Power Renew Energy Syst. 2014c; 1:36-46.
8. Bisoniya TS, Kumar A, Baredar P. Energy metrics of earth–air heat exchanger system for hot and dry climatic conditions of India. Energy Build. 2015; 86:214-221.
9. Bojic M, Trifunovic N, Papadakis G, Kyritsis S. Numerical simulation, technical and economic evaluation of air-to-earth heat exchanger coupled to a building. Energy. 1997; 22(12):1151-1158.

Veterinary Research Communications
https://doi.org/10.1007/s11259-022-09932-5

ORIGINAL ARTICLE

Effect of leaves and seeds of *Achyranthes aspera* as feed supplements on the immunological and stress parameters and related gene expressions of Asian catfish (*Clarias batrachus*)

Neelesh Kumar[1] · JaiGopal Sharma[1] · Prabhat Mittal[2] · Rina Chakrabarti[3]

Received: 13 January 2022 / Accepted: 27 April 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
The immunostimularory properties of *Achyranthes aspera* leaves and seeds supplemented feeds were evaluated in Asian catfish, *Clarias batrachus* (0.352±0.008 g). The experimental feeds contained 0.25% leaves (EFL1), 0.5% leaves (EFL2), 0.5% seeds (EFS) and 0% leaves or seeds (control feed; CF). In CF, leaves and seeds were absent. Fish were immunized with chicken-RBC after 60 days of feeding. The blood and tissue samples were collected on 7th, 14th and 21st days after immunization for various assays. The average weight of magur was significantly higher in EFS compared to the other treatments throughout the study period and EFL2 followed this group. Serum lysozyme level of fish was significantly higher in EFS on 7th and 14th days and in EFS and EFL2 on 21st day after immunization compared to the other treatments. Myeloperoxidase and nitric oxide synthase levels were always significantly higher in EFS diet fed fish compared to other treatments. The highest hemagglutination titer level was found in EFS throughout the study period. Thiobarbituric acid reactive substances (TBARS) and carbonyl protein levels were significantly lower in liver and kidney of enriched diets fed magur compared to CF treatment. TBARS and carbonyl protein levels were minimum in EFS diets fed fish. In EFS and EFL2 treatments, the expressions of *TNF-α*, *iNOS* and *NF-kB* were significantly higher compared to the CF group. *A. aspera* seeds and leaves showed significant immunostimulatory properties in Asian catfish fry.

Keywords *Achyranthes aspera* · *Clarias batrachus* · Lysozyme · Myeloperoxidase · *TNF-α* · *iNOS* and *NF-kB*

Introduction

The occurrence of infectious diseases has been affecting the aquaculture industry (Iwashita et al. 2015). In intensive aquaculture, the stressful condition weakened the immune system of fish and makes the organisms more susceptible to pathogen (Cerezuela et al. 2012; Giri et al. 2014). Therefore, it is most essential to improve the immunity of fish, so that the organism is able to overcome any stressful conditions prevailed in the aquatic environment. The non-specific and specific immune systems assist the fish to encounter pathogenic infection (Kumari and Sahoo 2005; Eslamloo et al. 2012). The non-specific immune response acts as a first line of defense; it represents a wide range of immune responses (Dalmo et al. 1997). Immunostimulants are a group of natural and synthetic compounds that increase the non-specific cellular and humoral defense responses. The boosting of immune system through medicinal plants is an eco-friendly and cost-effective approach. These plants are used as both chemotherapeutics and feed additives (Dawood et al. 2018). The immunostimulatory property of the plant is dose-dependent and the route of administration (*viz.*, injection, immersion, oral etc.) also plays important role (Awad and Awaad 2017). The supplementation of plant-based immunostimulant in diets improves the immune system of fish (Sharma et al. 2019, 2021).

Nitric oxide synthases (NOSs) are a family of enzymes responsible for catalyzing the synthesis of cellular signaling molecule nitric oxide (NO) from L-arginine. The inducible form (iNOS) is expressed in macrophages, Kupffer

✉ JaiGopal Sharma
sharmajaigopal@yahoo.com

[1] Department of Biotechnology, Delhi Technological University, Bawana Road, Delhi 110 042, India

[2] Satyawati College (Evening), University of Delhi, Ashok Vihar, Phase-III, Delhi 110 052, India

[3] Aqua Research Lab, Department of Zoology, University of Delhi, Delhi 110 007, India

Published online: 04 May 2022

cells, neutrophils, fibroblasts vascular smooth muscle and endothelial cells in response to pathogens (Eddy 2005). NO is a key mediator of innate (non-specific) host defense mechanism and involved in a variety of responses to infection and tissue damage. In addition, NOS activates cyclic guanosine monophosphate (cGMP), which induces mitochondrial KATP channel opening (Mattila and Thomas 2014; Maksin-Matveev et al. 2015). The tumor necrosis factor-α (*TNF-α*) is a proinflammatory cytokine and also one of the early immune genes which is expressed at early stage of infection in fish. It has an important role in regulation of inflammation. *TNF-α* triggers the expression of various immune genes (*IL-1β, IL-8, COX-2, IL-17C* etc.) which participate in inflammation (Rashmeei et al. 2020; Shekarabi et al. 2021). The *NF-kB* pathway was also activated due to stimulatory effects of *TNF-α* (Zhang et al. 2012; Hong et al. 2013). The TLR signaling pathway via activation of *NF-kB* gene, induces proinflammatory cytokines such as interleukin, type I interferon molecules, tumor necrosis factor molecules that provide direct defense response as well as alert adaptive immune cells against pathogens (Iwasaki and Medzhitov 2004; Kawai and Akira 2010). The study of some immunological parameters *viz*., lysozyme, myeloperoxidase, nitric oxide synthase and expression patterns of immune-related genes help to understand the capacity of fish to overcome the environmental stress and diseases (Kumar et al. 2019; Sharma et al. 2022).

Achyranthes aspera L. (Amaranthaceae), prickly chaff flower is commonly found as a weed in India (Saini et al. 2016) and in different tropical countries. Various parts of the plant are used for the treatment of human diseases like, inflammation, diabetes, hypertension, pneumonia, diarrhea, dysentery, asthma etc. (Baraik et al. 2014). The immunostimulatory and disease resistance properties of seeds and leaves of *A. aspera* have been evaluated in carps like rohu *Labeo rohita*, catla *Catla catla*, and in common carp *Cyprinus carpio* (Rao and Chakrabarti 2005; Rao et al. 2006; Chakrabarti and Srivastava 2012; Chakrabarti et al. 2012; Sharma et al. 2019; Singh et al. 2019).

Asian catfish *Clarias batrachus* (Linn.), popularly called magur is a bottom dwelling, carnivore, freshwater fish. It belongs to the family Clariidae (Roy et al. 2019). It is one of the most widespread catfish genera in the world (Thomas et al. 2013). The fish is marketed in living condition; it has high nutritional value and consumer preference (Paul et al. 2015). *C. batrachus* has been considered as critically endangered due to its habitat destruction and over-exploitation (Binoy 2010). Many studies show the outbreak of diseases in magur, like ulcerative disease, furunculosis and fungal infection (Thomas et al. 2013; Chattopadhyay et al. 1992; Chauhan 2012; Sarkar and Rashid 2012). Therefore, it is essential to improve the immunity of this economically important fish using natural immunostimulants like, leaves and seeds of *A. aspera*. There is scarcity of literature in the field of evaluation of immunostimulation of Asian catfish using herbal ingredients.

The present investigation aims to study the effect of feeding of *Achyranthes aspera* leaves and seeds enriched diets in strengthening the immune system and reduction of oxidative stress (with special reference to lipid and protein oxidations) in Asian catfish *Clarias batrachus*.

Materials and methods

Preparation of feed, culture of fish and immunization

The feeds used for the study were, experimental feed containing 0.25% leaves (EFL1), experimental feed containing 0.5% leaves (EFL2), experimental feed containing 0.5% seeds (EFS) of *Achyranthes aspera* and control feed containing 0% leaves/seeds (CF). *A. aspera* plants were grown in the outdoor facility. Leaves and ripe seeds were collected, cleaned and ground. The powder of leaves and seeds were added in the experimental feeds. The dry fish powder, wheat flour, cod liver oil and vitamin-mineral premix were other ingredients used for the formulation of both control and experimental feeds (Table 1). The doses of seeds and leaves of *A. aspera* were selected based on the earlier study (Rao et al. 2006; Chakrabarti and Srivastava 2012; Singh et al. 2020).

Clarias batrachus were obtained from a fish farm and randomly distributed in twelve tanks at a stocking density of 12 fish/tank (60 L). Fish were fed with control feed prior to the start of the experiment (Kumar et al. 2019). After 7 days of acclimation, fish were looking healthy and were accepting food properly; then the feeding trial started. The fry (0.352 ± 0.008 g) were fed with experimental and control feeds. Three replicates were used for each feeding scheme (36 fish/treatment). A separate group of fish was fed with control feed and they served as sham control during immunization. The feeding rate was 5% of the body weight of the fish. The excess feed was siphoned after 1 h of feeding, oven dried and measured. The guidelines of Animal Ethics Committee, Department of Zoology, University of Delhi (DU/ZOOL/IAEC-R/2015/08) was followed for the culture of fish and sampling.

The chicken blood was collected in Alsever's solution (1:3) and stored at 4°C for 24 h. The settled chicken (c-RBC) was washed thrice with phosphate buffer saline (1X PBS, pH 7.4). Then 80 mL of 1X PBS (pH 7.4) was mixed with 20 mL of c-RBC. This solution (20%, v/v) was used for the immunization of fish. After 60 days of feeding, the fish increased in size. Fish from all culture tanks were collected and anaesthetized with tricaine methanesulfonate (MS222,

Table 1 Formulation of fish diets and evaluation of their proximate composition

Ingredient (g/kg diet)	Diets[1]			
	CF	EFL1	EFL2	EFS
Feed composition				
Fishmeal	651.6	651.6	651.6	651.6
Wheat flour	334.4	331.9	329.4	329.4
Cod liver oil	10.0	10.0	10.0	10.0
[2]Vitamin and mineral complex	4.0	4.0	4.0	4.0
Leaves powder	-	2.5	5.00	-
Seeds powder	-	-	-	5.0
Proximate composition (g/100 g on dry matter basis)				
Moisture	7.14	11.09	10.03	9.43
Crude protein	36.96	37.05	36.98	37.02
Crude fat	8.64	9.24	7.82	8.31
Total carbohydrates	40.06	34.47	37.05	37.21
Ash	7.20	8.15	8.12	8.03
Crude fibre	4.1	6.8	6.5	4.4
[3]Energy (kcal/100 g)	385.84	369.24	366.50	371.71

[1]CF = Control feed, EFL1 = 0.25% leaves supplemented feed, EFL2 = 0.5% leaves supplemented feed,

EFS = 0.5% seeds supplemented feed

[2]Supradyan multivitamin tablets with minerals and trace elements contains (as mg/kg in diets): = Vitamin A (as acetate) 12; Cholecalciferol 0.1; Thiamine mononitrate, 40; Riboflavine 40; Pyridoxine hydrochloride, 12; Cyanocobalamin, 0.06; Nicotinamide, 400; Calcium pantothenate 65.20; Ascorbic acid 600; α-Tocopheryl acetate, 100; Biotin, 1.00. Minerals: Tribasic calcium phosphate, 516; Magnesium oxide, 240; Dried ferrous sulphate, 128.16; Manganse sulphate monohydrate 8.12; Total phosphorus, 103.20. Trace elements: Copper sulphate pentahydrate 13.56; Zinc sulphate, 8.80; Sodium molybdate dihydrate, 1.00; Sodium borate 3.52

[3]Energy (kcal/kg) = [(Crude protein g/kg × 4) + (Crude lipid g/kg × 9) + (Total Carbohydrate g/kg × 4)]

Sigma, USA) at the concentration of 25 mg/L of water (Carter et al. 2011). Then the weight of individual fish was measured and fish were immunized (the day of immunization, 0 day). The fish was injected intraperitoneally with chicken-RBC (50 µL/fish) using 26G × ½ syringe (0.45 × 13 mm). In earlier study, c-RBC was used for the immunization of fish (Rao et al. 2004) and same protocol was followed in the present study. Five control diet fed fish were injected with phosphate buffer saline (PBS) to check the effect of handling stress and they served as the sham control. Fish were introduced into the respective tank after immunization. Then magur were collected on 7th, 14th and 21st days after immunization. Twelve fish per treatment were sampled on each day. Water temperature (25.6 - 26.5°C), pH (7.5 - 7.67) and dissolved oxygen levels (5.10 - 6.25 mg/L) of culture tanks were recorded at weekly interval using HACH (USA) water quality analyzer. The duration of the study was 81 days.

The weight of all fish (36 fish/treatment) was measured on the day of immunization (0 day) and on 7th, 14th and 21st days (12 fish/treatment) after immunization using electronic balance (Shimadzu, AUX220, Japan).

Specific growth rate (SGR) of fish was calculated using the formula:

$$\text{SGR}\ (\%) = \frac{(\text{In final weight of fish} - \text{In initial weight of fish})}{\text{Duration of experiment}} \times 100$$

Feed conversion ratio (FCR) was calculated using the formula:

$$\text{FCR} = \frac{\text{Dry weight of food consumed by individual fish during experiment}}{\text{Wet weight gain of individual fish}}$$

Collection of samples and assays

On 7th day after immunization, fish were collected (12 fish/treatment), anaesthetized with MS222 and samples were collected for various assays. The blood sample was collected from the caudal vein of individual fish using sterilized syringe (1 mL) fitted with 26G × ½ "(0.45 × 13 mm) needle (Rao et al. 2004). The syringe and needle were rinsed with 2.7% ethylenediaminetetraacetic acid (EDTA) solution. The blood samples were allowed to clot in a sterilized microcentrifuge tube (1.5 mL) and stored in a refrigerator (4°C) overnight. After 12 h sample was centrifuged at 400 × *g* for 10 min at 4°C and serum was collected in sterile tube and stored at -20°C for the assay of various immunological enzymes. The liver and kidney were collected from each treatment (4 fish/treatment) and stored at -80°C for various enzymatic assays. Similarly, tissues of 4 fish per treatment were kept in TRIzol® reagent (Thermo Scientific, USA) for gene expression study. Like 7th day, samples were also collected on 14th and 21st days of immunization from each tank.

The Microplate Reader (Synergy H1 Hybrid Reader, BioTek, USA) was used for the assay of various enzymes. The serum lysozyme was measured using *Micrococcus lysodeikticus* (20 mg/100 mL sodium acetate buffer 0.02M., pH 5.5) solution (Siwicki 1989). Serum (10 µL) was mixed with 100 µL of solution and the initial absorbance was recorded at 450 nm. The final absorbance was measured after 1 h incubation. The lysozyme level was presented as mg/mL.

Myeloperoxidase levels of fish in various culture conditions were measured (Quade and Roth 1997). In each well of a microplate, 10 µL serum, 90 µL Hank's balanced salt solution and 35 µL of TMB/H_2O_2 were added, properly mixed and allow to incubate. After 2 min of incubation, 35 µL of H_2SO_4 was added to stop the reaction. Hanks' Balanced Salt Solution (HBSS) was used as blank. The standard curve

was prepared with 1,3,5-trinitrobenzene (TNB). The enzyme activity was estimated at 450 nm.

The hemagglutination assay was performed to measure the antigen-specific antibody response (Singh 1983; Das et al. 2000). Chicken blood (c-RBC) was collected in Alsever's solution (1:3) and stored at 4°C for 24 h. Chicken-RBC was washed thrice with PBS (pH 7.4) and centrifuged, then mixed in the PBS to make 2% (v/v) solution. A 50 µL serum was serially diluted with PBS in a 96-well round bottom ELISA plate. A 50 µL of c-RBC (2%) was added to each well, incubated for 60 min at room temperature (25°C), and then overnight at 4°C. The reciprocal of the highest dilution that represented agglutination was considered as the hemagglutination antibody titer.

The nitric oxide synthase levels of liver and kidney of magur was assayed (Lee et al. 2003). Tissue (100 mg) was homogenized in 1x PBS (pH 7.4) and centrifuged at 10,000 × *g* at 4°C for 15 min. The supernatant (100 µL) was mixed with 100 µL of Griess reagent (1% sulfanilamide and 0.1% naphthyl ethylenediamine in 5% phosphoric acid) and incubated at room temperature for 10 min. The absorbance was recorded at 540 nm. Distilled water was used as blank. The standard curve was prepared with sodium nitrite. The concentration of nitric oxide was expressed in µM/ mg tissue.

The measurement of thiobarbituric acid reactive substances (TBARS) indicates the tissue lipid oxidation. The liver/kidney was homogenized in KCl solution (1 g in 900 µL). The assay mixture was composed of the following: 25 µL homogenate, sodium dodecyl sulphate (25 µL), acetic acid (187.5 µL), thiobarbituric acid (187.5 µL) and distilled water (75 µL). After 1 h of incubation in water bath (95°C), the sample was brought into room temperature and a mixture of distilled water (125 µL) and *n*-butanol-pyridine (625 µL) was added to the sample. The whole reaction mixture was shaken vigorously and centrifuged. Then TBARS level (µmol malondialdehyde/mg tissue) of the sample was measured (532 nm). The standard curve was prepared with 1, 1, 3, 3-tetramethoxy propane (Ohkawa et al. 1979).

The carbonyl protein levels in liver and kidney were assayed (Lenz et al. 1989). The tissue (100 mg) was homogenized in 1 mL of 50 mmol potassium phosphate buffer (containing ethylenediaminetetraacetic acid and phenylmethylsulfonyl fluoride). The homogenate (250 µL) was mixed with 0.5 mL of trichloro acetic acid (10%) and centrifuged at 13000 x *g* for 5 min. The supernatant was removed and the pellet was mixed with 1 mL of 10 mM dinitrophenyl hydrazine. The mixture was kept at room temperature for 1 h. and centrifuged at 13000 x *g* for 5 min. The pellet was washed thrice with 1 mL of ethanol-butylacetate (1/1, v/v). The pellet was completely dissolved with 6 M guanidine hydrochloride and again centrifuged at 13000 x *g* for 5 min. The absorbance of supernatant was recorded at 370 nm. The carbonyl protein was expressed as nmol/mg protein. The molar extinction coefficient was 22×10^3/M/cm. Total protein content was estimated (750 nm) using bovine serum albumin as standard (Lowry et al. 1951). Three replicates were used for each assay.

Relative mRNA expression study

The liver and kidney of magur were kept in TRIzol reagent (Invitrogen, ThermoFisher Scientific, USA) at 4°C for the study of relative mRNA expressions of different genes *viz*., tumor necrosis factor-α (*TNF-α*), nitric oxide synthase (*iNOS*), nuclear factor-kB (*NF-kB*). The RNA was extracted from the tissue following manufacturer's protocol. Briefly, the preserved sample (100 mg) was taken out and was incubated at 25°C for 15 min after proper mixing. Then 200 µL chloroform (Sigma-Aldrich, USA) was added to the sample and was vigorously shaken for 30 sec and incubated at 25°C for 10 min. After incubation, the sample was centrifuged at 12000 x *g* at 4°C for 15 min. The upper aqueous layer was transferred into a 1.5 mL RNase free microcentrifuge tube containing 1 µL of glycogen and 500 µL 2-propanol/ isopropanol (Sigma-Aldrich) was added gently to the sample and incubated at 25°C for 10 min. The sample was centrifuged at 12000 x *g* at 4°C for 10 min and supernatant was removed. Pellet was mixed with 1 mL of 75% ethanol (Merck, USA) and centrifuged at 7500 x *g* at 4°C for 5 min. The supernatant was removed and the pellet was air dried till it became transparent. Finally, it was mixed with 40 µL of deionized, diethylpyrocarbonate (DEPC)-treated water (ThermoFisher Scientific, USA) and incubated at 58°C in an oven for 10 min and stored at -20°C for further use.

The NanoDrop spectrophotometer (Thermo Scientific, USA) was used to check the purity of RNA (260:280). The quality of RNA was tested in gel electrophoresis (agarose 1%). The RNA (1 µg) was first treated with 1U of DNase to avoid contamination of DNA and then used for cDNA preparation through reverse transcription with random primers. The High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, ThermoFisher Scientific) was used for the study. The cDNA synthesis confirmation was checked with housekeeping gene (*β-actin*) through PCR-amplification.

The relative quantification of various genes was performed with PowerUp™ SYBR™ Green Master Mix (Applied Biosystems, ThermoFisher Scientific). The real-time PCR (Applied Biosystems, QuantStudio 6 Flex, USA) was used for the estimation. The primers for specific genes were designed using NCBI tool (Table 2). The efficiency and specificity of primers were also checked using a serial dilution of cDNA and constructing a standard curve with correlation coefficient. The reaction mixture (10 µL) consisted of cDNA (1 µL, 1:3), 2 × PowerUp™ SYBR™ Green PCR Master Mix (5 µL), 2.5 µmol forward

Table 2 The primers of the target genes and their sequences used for qPCR analysis

Target gene primers	Forward/ reverse	Nucleotide sequence (5' - 3')	T_m (^{0}C)	Product length (bp)	Accession number
TNF-α	Forward reverse	TCTCAGGTCAATACAACCCGC GAGGCCTTTGCGGAAAATCTTG	60 ^{0}C	125	KM593875
iNOS	Forward reverse	ATGGGCACTGAGATTGGAGC CTTCGTCCTTCCACAGCGAT	55 ^{0}C	121	KT180212
NF-kB	Forward reverse	CTCGGTGAAGCGAGGAAGAG CATTCCGGCTCCGACTCTC	60 ^{0}C	124	MG571500
β-actin	Forward reverse	GAGCACCCTGTCCTGCTTAC GTACAGGGACAGCACAGCC	60 ^{0}C	132	EU527190

TNF-α: Tumour necrosis factor α; *iNOS: Inducible nitric oxide synthase*; *NF-kB: Nuclear factor kappa B*; *β-actin: Beta-actin*

and reverse primers (0.25 µL each) and nuclease free water (3.5 µL). The conditions of real-time PCR were as follows: 95°C for 10 min, 40 cycles at 95°C for 15 sec and 60°C for 1 min. All samples were run in duplicate and a non-template control (NTC) was also run without cDNA. The specificity of qPCR was checked through melt curve analysis. The conditions for melt curve analysis are as follows: 95°C for 15 sec, 60°C for 1 min, and at 95°C for 15 sec. After completing the programme, the threshold cycle (C_T) value of each well was recorded and analysis was performed using standard $2^{-\Delta\Delta C}{}_T$ method (Livak and Schmittgen 2001). It was considered that primers were amplified with 100% efficiency. The *β-actin* was used as a housekeeping gene and also used for normalization of target genes. The data of treatment groups were compared to the control group.

Statistical analysis

The data for growth parameters, various enzymes and mRNA expressions are provided as means ± standard error. The Kolmogorov-Smirnov test was used to ascertain the normality of the data of average weight of fish and of various parameters in different treatments. The results indicated normality in the data ($p > 0.05$), except for one parameter hemagglutination titer, which resulted in significance value $p < 0.05$ in all treatment categories. The one way ANOVA analysis was performed among four different treatments and also for three different days of sampling (7, 14 and 21) for each parameter in the treatment groups. The non-parametric Kruskal-wallis test was applied to study the differences in hemagglutination titer among different treatments and different days. Further, Tukey HSD post-hoc analysis confirmed the pair-wise significance among different treatments and different days of individual treatment for each parameter. Statistical analysis was performed using SPSS 26.0 version (South Asia (P) Ltd., Bangalore, India) software. The significance level was accepted as $p < 0.05$.

Results

Growth performance

The average weight of magur was significantly higher in EFS compared to other three treatments throughout the experimental period (Table 3). Highest growth of fish was found on 21st day after immunization regardless of feeding schemes. The SGR was significantly higher in EFS

Table 3 Average weight, specific growth rate and feed conversion ratio found in Asian catfish (*Clarias batrachus*) cultured under different feeding regime. Means with different superscripts in the same row are significantly different (n=3)

Samplings	Experimental diets			
	CF	EFL1	EFL2	EFS
Final average weight (g)				
0th Day	2.23 ± 0.14^c	2.37 ± 0.13^c	3.08 ± 0.17^b	3.88 ± 0.20^a
7th Day	3.01 ± 0.11^d	3.68 ± 0.18^c	4.35 ± 0.16^b	4.95 ± 0.14^a
14th Day	3.69 ± 0.12^d	4.15 ± 0.09^c	5.03 ± 0.23^b	7.08 ± 0.21^a
21st Day	4.10 ± 0.09^c	5.83 ± 0.30^b	5.91 ± 0.22^b	7.79 ± 0.30^a
Specific growth rate (%)				
0th Day	3.07 ± 0.11^c	3.17 ± 0.09^c	3.61 ± 0.09^b	4.00 ± 0.09^a
7th Day	4.35 ± 0.63^c	6.30 ± 0.37^a	4.98 ± 0.53^b	3.52 ± 0.52^d
14th Day	2.90 ± 0.04^b	1.75 ± 0.23^d	2.07 ± 0.14^c	5.10 ± 0.05^a
21st Day	1.52 ± 0.09^c	4.84 ± 0.25^a	2.30 ± 0.07^b	1.37 ± 0.08^c
Feed conversion ratio (FCR)				
0th Day	0.46 ± 0.04^a	0.42 ± 0.03^a	0.31 ± 0.02^b	0.24 ± 0.02^c
7th Day	0.83 ± 0.14^b	0.51 ± 0.04^d	0.69 ± 0.09^c	1.05 ± 0.18^a
14th Day	1.25 ± 0.02^c	2.22 ± 0.29^a	1.81 ± 0.13^b	0.65 ± 0.01^d
21st Day	2.51 ± 0.16^b	0.70 ± 0.04^d	1.60 ± 0.06^c	2.80 ± 0.16^a

CF = Control feed, EFL1 = 0.25% leaves supplemented feed, EFL2 = 0.5% leaves supplemented feed, EFS = 0.5% seeds supplemented feed

compared to other treatments on the day of immunization and on 14th day after immunization. The FCR was significantly lower in EFS compared to other treatments on the day of immunization and on 14th day after immunization.

Biochemical assays

Serum lysozyme, myeloperoxidase and hemagglutination titer

Serum lysozyme, myeloperoxidase and hemagglutination titer levels of magur cultured in four treatments were estimated on three different days after immunization (Table 4). The serum lysozyme level was significantly higher in EFS compared to other treatments on 7th and 14th days after immunization. The levels were significantly higher in EFS and EFL2 compared to other two feeding regimes on 21st day after immunization. The myeloperoxidase level was significantly higher EFS compared to other treatments in all days of sampling. The hemagglutination titer level was maximum in EFS throughout the study period.

The days of immunization also influenced these parameters. Serum lysozyme and myeloperoxidase levels were significantly higher on 21st day after immunization regardless of treatments, except myeloperoxidase level in the control diet fed (CF) magur. In the CF, myeloperoxidase level was maximum on 14th day after immunization. In all these treatments, highest hemagglutination titer level was found on 14th day after immunization.

Nitric oxide synthase, thiobarbituric acid reactive substances and carbonyl protein

The nitric oxide synthase level in liver of EFS diet fed magur was significantly higher compared to other treatments throughout the study period (Table 5). The nitric oxide synthase level was significantly higher in kidney of EFS diet fed magur. The TBARS level was significantly lower in liver of EFS diet fed fish compared to the fish fed other diets in all days of sampling. Similar trend was also found in kidney. In both liver and kidney of EFS diet fed magur, the carbonyl protein levels were significantly lower compared to other treatments throughout the study period.

The effect of days of immunization was also found in this study. The nitric oxide synthase levels in liver and kidney increased gradually with the days of immunization. Highest nitric oxide synthase level was found on 21st day after immunization in all these treatments. TBARS and carbonyl protein showed the opposite trends. In liver and kidney, TBARS levels gradually decreased with the days of immunization regardless of treatments. Similar trend was found with carbonyl protein. In all these feeding schemes, carbonyl levels were significantly higher on 7th day and lower on 21st day after immunization compared to other days of sampling.

Relative mRNA expression study

In liver and kidney of magur, *TNF-α, iNOS* and *NF-kB* genes expressions were significantly higher in EFS compared to other treatment throughout the study period (Fig. 1). Treatment EFL2 followed the EFS. In EFL1, *NF-kB* was down-regulated in kidney on 7th and 14th days after immunization compared to the control diet fed fish. There were

Table 4 Serum lysozyme (mg/mL), myeloperoxidase (O.D. at 450 nm) and hemagglutination titer levels found in Asian catfish (*Clarias batrachus*). Means with different superscripts in the same row are significantly different (n=3)

Parameters	Experimental diets			
	CF	EFL1	EFL2	EFS
7 th day				
Serum lysozyme	57.98 ± 3.09^{d}	63.03 ± 1.17^{c}	80.04 ± 3.59^{b}	88.21 ± 3.21^{a}
Myeloperoxidase	1.48 ± 0.01^{d}	1.61 ± 0.04^{c}	1.79 ± 0.02^{b}	1.93 ± 0.04^{a}
Hemagglutination	4.5 ± 1.26^{d}	5 ± 1.00^{c}	7 ± 1.00^{b}	10 ± 1.15^{a}
14 th day				
Serum lysozyme	50.29 ± 8.49^{d}	74.14 ± 5.55^{c}	203.93 ± 1.88^{b}	218.93 ± 8.44^{a}
Myeloperoxidase	1.57 ± 0.01^{d}	1.78 ± 0.02^{c}	1.93 ± 0.03^{b}	2.23 ± 0.04^{a}
Hemagglutination	10 ± 2.00^{c}	10 ± 2.00^{c}	14 ± 2.00^{b}	24 ± 4.62^{a}
21st day				
Serum lysozyme	97.86 ± 3.80^{c}	129.29 ± 8.01^{b}	279.05 ± 9.30^{a}	282.50 ± 5.96^{a}
Myeloperoxidase	1.25 ± 0.02^{d}	1.85 ± 0.04^{c}	2.17 ± 0.03^{b}	2.52 ± 0.05^{a}
Hemagglutination	6 ± 1.15^{b}	7 ± 1.00^{b}	10 ± 2.00^{a}	14 ± 2.00^{a}

CF = Control feed, EFL1 = 0.25% leaves supplemented feed, EFL2 = 0.5% leaves supplemented feed, EFS = 0.5% seeds supplemented feed

Table 5 Nitric oxide synthase, NOS (µmol/mg tissue), thiobarbituric acid reactive substances, TBARS (µmol MDA/mg protein) and carbonyl protein (nmol/mg protein) levels found in Asian catfish (*Clarias batrachus*). Means with different superscripts in the same row are significantly different (n=3)

Tissues	Parameters	Experimental diets			
		CF	EFL1	EFL2	EFS
7th day					
Liver	NOS	0.68 ± 0.04^{d}	1.10 ± 0.01^{c}	1.72 ± 0.03^{b}	2.38 ± 0.04^{a}
	TBARS	9.30 ± 0.27^{a}	4.70 ± 0.08^{b}	2.54 ± 0.14^{c}	1.30 ± 0.02^{d}
	Carbonyl protein	21.55 ± 0.43^{a}	9.85 ± 0.22^{b}	6.69 ± 0.22^{c}	4.69 ± 0.07^{d}
Kidney	NOS	0.11 ± 0.02^{d}	0.45 ± 0.03^{c}	0.96 ± 0.02^{b}	1.13 ± 0.05^{a}
	TBARS	11.34 ± 0.33^{a}	6.91 ± 0.06^{b}	4.11 ± 0.07^{c}	1.87 ± 0.08^{d}
	Carbonyl protein	21.02 ± 1.35^{a}	11.17 ± 0.18^{b}	7.55 ± 0.18^{c}	3.40 ± 0.12^{d}
14 th day					
Liver	NOS	0.94 ± 0.03^{d}	1.77 ± 0.01^{c}	2.39 ± 0.02^{b}	3.06 ± 0.04^{a}
	TBARS	7.02 ± 0.18^{a}	4.06 ± 0.09^{b}	1.97 ± 0.04^{c}	1.16 ± 0.04^{d}
	Carbonyl protein	14.43 ± 0.27^{a}	7.96 ± 0.33^{b}	5.32 ± 0.07^{c}	3.92 ± 0.11^{d}
Kidney	NOS	0.27 ± 0.03^{d}	0.74 ± 0.02^{c}	1.15 ± 0.03^{b}	1.37 ± 0.04^{a}
	TBARS	9.58 ± 0.10^{a}	5.05 ± 0.09^{b}	2.78 ± 0.03^{c}	1.27 ± 0.02^{d}
	Carbonyl protein	17.96 ± 0.28^{a}	7.60 ± 0.22^{b}	4.72 ± 0.11^{c}	2.22 ± 0.11^{d}
21st day					
Liver	NOS	1.53 ± 0.03^{d}	2.47 ± 0.04^{c}	3.15 ± 0.03^{b}	$3.76 \pm 0.04a$
	TBARS	5.92 ± 0.21^{a}	3.72 ± 0.07^{b}	1.62 ± 0.05^{c}	0.85 ± 0.03^{d}
	Carbonyl protein	12.27 ± 0.21^{a}	6.04 ± 0.25^{b}	2.38 ± 0.21^{c}	1.13 ± 0.14^{d}
Experimental diets	NOS	0.42 ± 0.03^{d}	0.94 ± 0.02^{c}	1.31 ± 0.02^{b}	1.57 ± 0.02^{a}
	TBARS	8.51 ± 0.19^{a}	4.10 ± 0.05^{b}	2.14 ± 0.07^{c}	0.93 ± 0.04^{d}
	Carbonyl protein	14.70 ± 0.29^{a}	5.82 ± 0.07^{b}	3.48 ± 0.14^{c}	1.12 ± 0.05^{d}

CF = Control feed, EFL1 = 0.25% leaves supplemented feed, EFL2 = 0.5% leaves supplemented feed, EFS = 0.5% seeds supplemented feed

up-regulations of all these genes in EFL1 on 21st day after immunization.

Discussion

In magur, the effect of *A. aspera* seeds and leaves supplemented diets were recorded. The supplementation of leaves and seeds at 0.5% level (EFL2 and EFS) enhanced the growth of magur compared to the control (CF) diet fed fish. The enhanced growth was also found in rohu fed with *A. aspera* seeds and leaves (Sharma et al. 2019; Srivastava and Chakrabarti 2012) and guava *Psidium guajava* leaves (Giri et al. 2015) supplemented diets. The chemical composition study showed the nutritional value of seeds and leaves. The seeds contain all essential amino acids, fatty acids, different oleonic acid and ecdysterone (Varuna et al. 2010). The presence of ecdysterone influenced the growth of fish (Chakrabarti et al. 2012). All essential amino acids were present in the seeds and leaves. The amino acid contents were higher in seeds compared to leaves (Kumar et al. 2021). Various secondary metabolites like, alkaloids, flavonoids, saponins, steroids and terpenoids were reported in leaves (Umamaheswari et al. 2012) and their presence influenced the performance of magur.

The fatty acid profiles of seeds and leaves influenced the immune system of enriched diets fed fish. The monounsaturated fatty acids (MUFA) and n-6 polyunsaturated (PUFA) contents were higher in seeds compared to leaves, whereas n-3 PUFA contents were higher in leaves compared to the seeds (Kumar et al. 2021). The higher lysozyme levels were found in EFS and EFL2 diets fed fish compared to the other diets fed fish throughout the study period. Lysozyme modulates host immune system against infections (Ragland and Criss 2017). The myeloperoxidase level was always maximum in EFS. The lysosomal protein myeloperoxidase plays significant role in non-specific immune system (Hampton and Orrenius 1998). Minimum hemagglutination titer levels were observed in EFL1 and CF. Significantly higher levels of nitric oxide synthase were recorded in liver and kidney of EFS diet fed magur compared to others. The EFL2 diet fed magur followed this treatment. In earlier study, higher levels of lysozyme, myeloperoxidase and nitric oxide synthase were found in seeds and leaves supplemented diets fed catla and rohu (Sharma et al. 2019; Chakrabarti et al. 2014). The presence of long-chain polyunsaturated fatty acids such as linolenic and oleic acids in the seeds may be associated with the immunostimulatory properties (Chakrabarti et al. 2012).

The increased levels of TBARS and carbonyl protein in the tissue are indicators of oxidative damage (Parvez and

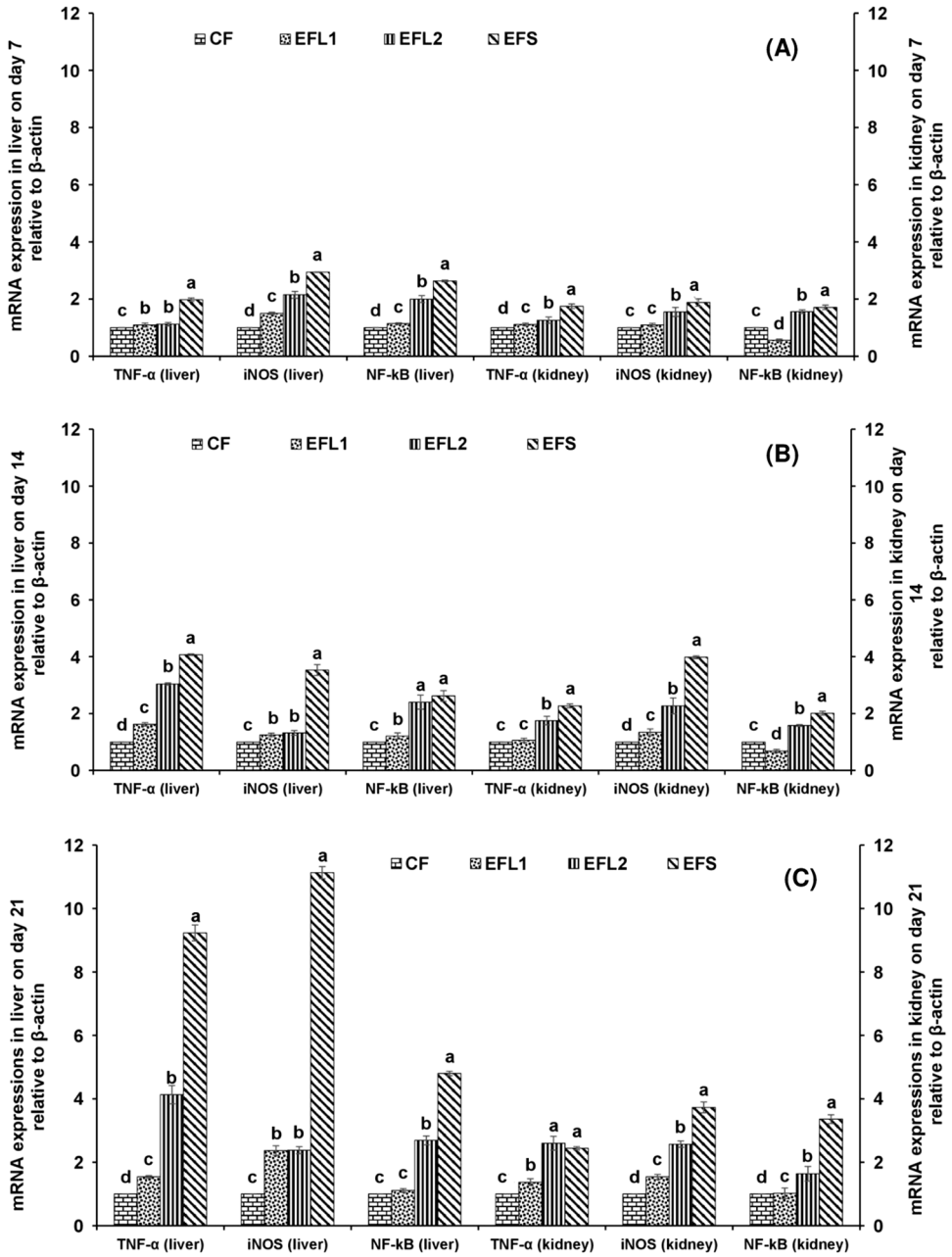


Fig. 1 Expression of *TNF-α*, *iNOS*, *NF-kB* in liver and kidney of four different diets fed Asian catfish (*Clarias batrachus*) on (A) 7th day, (B) 14 th day and (C) 21 st day after immunization. The relative expressions of the target genes were normalized to the expression of *β-actin* (internal control) and expressed as fold changes relative to the control diet fed group. Vertical bars with different superscripts are significantly ($p < 0.05$) different (n=3)

Raisuddin 2006). TBARS and carbonyl protein contents were always minimum in EFS diet fed magur. The presence of vitamin C in the seeds reduced the oxidative stress in the enriched diet fed fish (Kumar et al. 2021). Similar results were found in *A. aspera* seeds and leaves supplemented diets fed rohu (Sharma et al. 2019). The seeds supplemented diet gave protection against oxidative damage to UV-B radiated fish (Singh et al. 2013a; Singh et al. 2013b; Sharma et al. 2015).

Higher expressions of *TNF-α*, *iNOS* and *NF-kB* were found in liver and kidney of EFS and EFL2 diets fed magur compared to the control diet fed fish. The expressions were higher in EFS compared to EFL2. Higher expression of *TNF-α* was observed in seeds and leaves enriched diets fed rohu (Sharma et al. 2019). The up-regulation of *TNF-α* was found in guava leaves supplemented diet fed rohu (Giri et al. 2015). *TNF-α* supernatants treated endothelial cells promoted respiratory burst activity and migration of leucocytes in carps (Forlenza et al. 2009). The exposure of rohu at early stage to bisphenol-A caused oxidative stress and suppress *NF-κB* signaling pathway leading to immunosuppression (Faheem et al. 2020).

Conclusion

Achyranthes aspera seeds and leaves enriched diets showed immunostimulatory properties in fish. The seeds and leaves should be supplemented at 0.5% level in fish diets and seeds showed better result compared to leaves.

Acknowledgements We are thankful to Department of Biotechnology, DBT, Government of India, New Delhi, India (No. 102/IFD/SAN/4260/2014-2015) and Department of Science and Technology, Government of India, New Delhi, India (No. DST/INSPIRE Fellowship/2017/IF170573) have supported this study.

Availability of data and material The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Author contribution RC, JGS, PM: designed the study; NK, RC, JGS: cultured the fish and analyzed samples; RC, JGS, PM: prepared the manuscript; PM, RC: performed statistical analysis; NK, PM, RC: prepared graphs and tables.

Funding Department of Biotechnology, DBT, Government of India, New Delhi, India (No. 102/IFD/SAN/4260/2014-2015) and Department of Science and Technology, Government of India, New Delhi, India (No. DST/INSPIRE Fellowship/2017/IF170573) have supported this study.

Declarations

Ethics approval All experiments were performed following the guidelines and regulations with approval from the ethical committee of Institutional Animal Ethics Committee, IAEC, University of Delhi, Delhi, approval number DU/ZOOL/IAEC-R/2015/08 was followed for the culture of fish and sample analysis.

Consent to participate Not applicable.

Consent for publication All contributing authors have been asked to give consent to publication at the point of acceptance, to confirm that they have approved the final version of the manuscript and have made all required statements and declarations.

Conflict of interest No conflict of interest among authors.

References

Awad E, Awaad A (2017) Role of medicinal plants on growth performance and immune status in fish. Fish Shellfish Immunol 67:40–54. https://doi.org/10.1016/j.fsi.2017.05.034

Baraik B, Jain P, Sharma H (2014) *Achyranthes aspera* L.: as a source of bio-fungicide. Am J Adv Drug Deliv 2:686-696.

Binoy VV (2010) Catfish *Clarias* is vanishing from the waters of Kerala. Curr Sci 99:714

Carter KM, Woodley CM, Brown RS (2011) A review of tricaine methanesulfonate for anesthesia of fish. Rev Fish Biol Fish 21:51–59. https://doi.org/10.1007/s11160-010-9188-0

Cerezuela R, Guardiola FA, González P, Meseguer J, Esteban MÁ (2012) Effects of dietary *Bacillus subtilis*, *Tetraselmis chuii*, and *Phaeodactylum tricornutum*, singularly or in combination, on the immune response and disease resistance of sea bream (*Sparus aurata* L.). Fish Shellfish Immunol 33:342–349. https://doi.org/10.1016/j.fsi.2012.05.004

Chakrabarti R, Srivastava PK (2012) Effect of dietary supplementation with *Achyranthes aspera* seed on larval rohu *Labeo rohita* challenged with *Aeromonas hydrophila*. J Aquat Anim Health 24:213–218. https://doi.org/10.1080/08997659.2012.694834

Chakrabarti R, Srivastava PK, Kundu K, Khare RS, Banerjee S (2012) Evaluation of immunostimulatory and growth promoting effect of seed fractions of *Achyranthes aspera* in common carp *Cyprinus carpio* and identification of active constituents. Fish Shellfish Immunol 32:839–843. https://doi.org/10.1016/j.fsi.2012.02.006

Chakrabarti R, Srivastava PK, Verma N, Sharma JG (2014) Effect of seeds of *Achyranthes aspera* on the immune responses and expression of some immune-related genes in carp *Catla catla*. Fish Shellfish Immunol 41:64–69. https://doi.org/10.1016/j.fsi.2014.03.007

Chattopadhyay UK, Pal D, Das MS, Pramanik AK, Pal RN (1992) Association of fish and environmental strains of *Aeromonas hydrophila* in the causation of ulcerative disease syndrome in fish. Int J Environ Health Res 2:18–23. https://doi.org/10.1080/09603129209356727

Chauhan R (2012) Study on certain fungal diseases in culturable and nonculturable species of fishes of Upper Lake, Bhopal. J Chem Biol Phys 2:1810–1815

Dalmo RA, Ingebrigtsen K, Bøgwald J (1997) Non-specific defence mechanisms in fish, with particular reference to the reticuloendothelial system (RES). J Fish Dis 20:241–273. https://doi.org/10.1046/j.1365-2761.1997.00302.x

Das K, Singh BN, Pandey AK, Ayyappan S (2000) Immunological changes induced by ovaprim in the freshwater catfish, *Heteropneustes fossilis* (Bloch). Indian J Fish 47:123–127

Dawood MA, Koshio S, Esteban MÁ (2018) Beneficial roles of feed additives as immunostimulants in aquaculture: a review. Rev Aquac 10:950–974. https://doi.org/10.1111/raq.12209

Eddy FB (2005) Cardiac function in juvenile salmon (*Salmo salar* L) in response to lipopolysaccharide (LPS) and inhibitors of inducible nitric oxide synthase (iNOS). Fish Physiol Biochem 31:339–346. https://doi.org/10.1007/s10695-005-3528-3

Eslamloo K, Falahatkar B, Yokoyama S (2012) Effects of dietary bovine lactoferrin on growth, physiological performance, iron metabolism and non-specific immune responses of Siberian sturgeon *Acipenser baeri*. Fish Shellfish Immunol 32:976–985. https://doi.org/10.1016/j.fsi.2012.02.007

Faheem M, Adeel M, Khaliq S, Lone KP, El-Din-H-Sayed A (2020) Bisphenol-A induced antioxidants imbalance and cytokines alteration leading to immune suppression during larval development of *Labeo rohita*. Environ Sci Pollut Res 27:26800–26809. https://doi.org/10.1007/s11356-020-08959-y

Forlenza OV, Diniz BS, Talib LL, Mendonça VA, Ojopi EB, Gattaz WF, Teixeira AL (2009) Increased serum IL-1β level in Alzheimer's disease and mild cognitive impairment. Dement Geriatr Cogn Disord 28:507–512. https://doi.org/10.1159/000255051

Giri SS, Sukumaran V, Sen SS, Jena PK (2014) Effects of dietary supplementation of potential probiotic *Bacillus subtilis* VSG 1 singularly or in combination with *Lactobacillus plantarum* VSG 3 or/and *Pseudomonas aeruginosa* VSG 2 on the growth, immunity and disease resistance of *Labeo rohita*. Aquac Nutr 20:163–171. https://doi.org/10.1111/anu.12062

Giri SS, Sen SS, Chi C, Kim HJ, Yun S, Park SC, Sukumaran V (2015) Effect of guava leaves on the growth performance and cytokine gene expression of *Labeo rohita* and its susceptibility to *Aeromonas hydrophila* infection. Fish Shellfish Immunol 46:217–224. https://doi.org/10.1016/j.fsi.2015.05.051

Hampton MB, Orrenius S (1998) Redox regulation of apoptotic cell death in the immune system. Toxicol Lett 102:355–358. https://doi.org/10.1016/S0378-4274(98)00333-6

Hong S, Li R, Xu Q, Secombes CJ, Wang T (2013) Two types of *TNF-α* exist in teleost fish: phylogeny, expression, and bioactivity analysis of type-II *TNF-α3* in rainbow trout *Oncorhynchus mykiss*. J Immunol 191:5959–5972. https://doi.org/10.4049/jimmunol.1301584

Iwasaki A, Medzhitov R (2004) Toll-like receptor control of the adaptive immune responses. Nat Immunol 5:987–995. https://doi.org/10.1038/ni1112

Iwashita MKP, Nakandakare IB, Terhune JS, Wood T, Ranzani-Paiva MJT (2015) Dietary supplementation with *Bacillus subtilis*, *Saccharomyces cerevisiae* and *Aspergillus oryzae* enhance immunity and disease resistance against *Aeromonas hydrophila* and *Streptococcus iniae* infection in juvenile tilapia *Oreochromis niloticus*. Fish Shellfish Immunol 43:60–66. https://doi.org/10.1016/j.fsi.2014.12.008

Kawai T, Akira S (2010) The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. Nat Immunol 11:373–384. https://doi.org/10.1038/ni.1863

Kumar N, Sharma JG, Singh SP, Singh A, Krishna VH, Chakrabarti R (2019) Validation of growth enhancing, immunostimulatory and disease resistance properties of *Achyranthes aspera* in *Labeo rohita* fry in pond conditions. Heliyon 5:e01246. https://doi.org/10.1016/j.heliyon.2019.e01246

Kumar N, Sharma JG, Kumar G, Shrivastav AK, Tiwari N, Begum A, Chakrabarti R (2021) Evaluation of nutritional value of prickly chaff flower (*Achyranthes aspera*) as fish feed ingredient. Indian J Anim Sci 91:239–244

Kumari J, Sahoo PK (2005) Effects of cyclophosphamide on the immune system and disease resistance of Asian catfish *Clarias batrachus*. Fish Shellfish Immunol 19:307–316. https://doi.org/10.1016/j.fsi.2005.01.008

Lenz AG, Costabel U, Shaltiel S, Levine RL (1989) Determination of carbonyl groups in oxidatively modified proteins by reduction with tritiated sodium borohydride. Anal Biochem 177:419–425. https://doi.org/10.1016/0003-2697(89)90077-8

Lee DU, Kang YJ, Park MK, Lee YS, Seo HG, Kim TS, Kim CH, Chang KC (2003) Effects of 13-alkyl-substituted berberine alkaloids on the expression of COX-II, TNF-α, iNOS, and IL-12 production in LPS-stimulated macrophages. Life Sci 73:1401–1412. https://doi.org/10.1016/S0024-3205(03)00435-1

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. Methods 25:402–408. https://doi.org/10.1006/meth.2001.1262

Lowry OH, Rosebrough NJ, Farr AL, Randall RJ (1951) Protein measurement with the Folin phenol reagent. J Biol Chem 193:265–275. https://doi.org/10.1016/S0021-9258(19)52451-6

Maksin-Matveev A, Kanfi Y, Hochhauser E, Isak A, Cohen HY, Shainberg A (2015) Sirtuin 6 protects the heart from hypoxic damage. Exp Cell Res 330:81–90. https://doi.org/10.1016/j.yexcr.2014.07.013

Mattila JT, Thomas AC (2014) Nitric oxide synthase: non-canonical expression patterns. Front Immunol 5:478. https://doi.org/10.3389/fimmu.2014.00478

Ohkawa H, Ohishi N, Yagi K (1979) Assay for lipid peroxides in animal tissues by thiobarbituric acid reaction. Anal Biochem 95:351–358. https://doi.org/10.1016/0003-2697(79)90738-3

Parvez S, Raisuddin S (2006) Effects of paraquat on the freshwater fish *Channa punctata* (Bloch): non-enzymatic antioxidants as biomarkers of exposure. Arch Environ Contam Toxicol 50:392–397. https://doi.org/10.1007/s00244-005-5083-4

Paul P, Adikesavalu H, Banerjee S, Abraham TJ (2015) Antibiotic resistant motile aeromonads induced septicemia in Philippine catfish *Clarias batrachus* (Linnaeus, 1758) fingerlings. Croat J Fish 73:170-175. https://doi.org/10.14798/73.4.84

Quade MJ, Roth JA (1997) A rapid, direct assay to measure degranulation of bovine neutrophil primary granules. Vet Immunol Immunopathol 58:239–248. https://doi.org/10.1016/S0165-2427(97)00048-2

Ragland SA, Criss AK (2017) From bacterial killing to immune modulation: recent insights into the functions of lysozyme. PLoS Pathog 13:e1006512. https://doi.org/10.1371/journal.ppat.1006512

Rao YV, Chakrabarti R (2005) Stimulation of immunity in Indian major carp *Catla catla* with herbal feed ingredients. Fish Shellfish Immunol 18:327–334. https://doi.org/10.1016/j.fsi.2004.08.005

Rao YV, Romesh M, Singh A, Chakrabarti R (2004) Potentiation of antibody production in Indian major carp *Labeo rohita*, rohu, by *Achyranthes aspera* as a herbal feed ingredient. Aquaculture 238:67–73. https://doi.org/10.1016/j.aquaculture.2004.04.029

Rao YV, Das NK, Jyotyrmayee P, Chakrabarti R (2006) Effect of *Achyranthes aspera* on the immunity and survival of *Labeo rohita* infected with *Aeromonas hydrophila*. Fish Shellfish Immunol 20:263–273. https://doi.org/10.1016/j.fsi.2005.04.006

Rashmeei M, Shekarabi SPH, Mehrgan MS, Paknejad H (2020) Stimulatory effect of dietary chasteberry (*Vitex agnus-castus*) extract on immunity, some immune-related gene expression, and resistance against *Aeromonas hydrophila* infection in goldfish (*Carassius auratus*). Fish Shellfish Immunol 107:129–136. https://doi.org/10.1016/j.fsi.2020.09.037

Roy D, Mitra A, Biswas M, Chakraborty S, Pal S, Homechaudhuri S (2019) Early ontogeny of the Asian catfish magur, *Clarias batrachus* (Linnaeus, 1758). Int J Fish Aquat Stud 7:287–292

Saini RK, Sivanesan I, Keum YS (2016) Phytochemicals of *Moringa oleifera*: a review of their nutritional, therapeutic and industrial significance. 3 Biotech 6:1-14. https://doi.org/10.1007/s13205-016-0526-3

Sarkar MJA, Rashid MM (2012) Pathogenicity of the bacterial isolate *Aeromonas hydrophila* to catfishes, carps and perch. J Bangladesh Agril Univ 10:157–161. https://doi.org/10.22004/ag.econ.209312

Sharma JG, Singh MK, Chakrabarti R (2015) Physiological responses of *Catla catla* larvae fed with *Achyranthes aspera* seed enriched diet and exposed to UV-B radiation. Indian J Biochem Biophys 52:155–160

Sharma JG, Kumar N, Singh SP, Singh A, HariKrishna V, Chakrabarti R (2019) Evaluation of immunostimulatory properties of prickly chaff flower *Achyranthes aspera* in rohu *Labeo rohita* fry in pond conditions. Aquaculture 505:183–189. https://doi.org/10.1016/j.aquaculture.2019.02.065

Sharma JG, Singh A, Begum A, Krishna VH, Chakrabarti R (2021) The impact of *Achyranthes aspera* seeds and leaves supplemented feeds on the survival, growth, immune system and specific genes involved in immunostimulation in *Clarias batrachus* fry challenged with *Aeromonas hydrophila* in pond conditions. Fish Shellfish Immunol 118:11–18. https://doi.org/10.1016/j.fsi.2021.08.026

Sharma JG, Kumar N, Mittal P, Chakrabarti R (2022) Evaluation of UV–B protective properties of leaves and seeds of *Achyranthes aspera* in Asian catfish *Clarias batrachus* (Linn.). Photochem Photobiol Sci https://doi.org/10.1007/s43630-022-00222-2

Shekarabi SPH, Mostafavi ZS, Mehrgan MS, Islami HR (2021) Dietary supplementation with dandelion (*Taraxacum officinale*) flower extract provides immunostimulation and resistance against *Streptococcus iniae* infection in rainbow trout (*Oncorhynchus mykiss*). Fish Shellfish Immunol 118:180–187. https://doi.org/10.1016/j.fsi.2021.09.004

Singh AK (1983) Agglutination methods. In: Talwar GP (ed) A Handbook of Practical Immunology. Vikas Publishing, New Delhi, pp 139–149

Singh KM, Sharma JG, Chakrabarti R (2013a) Impact of UV-B radiation on the physiology of freshwater carp *Labeo rohita* larvae and evaluation of UV-B protective properties of seeds of *Achyranthes aspera* and vitamin C. Agric Res 2:166–171. https://doi.org/10.1007/s40003-013-0060-z

Singh MK, Sharma JG, Chakrabarti R (2013b) Effect of UV-B radiation on the defence system of *Labeo rohita* (Actinopterygii: Cypriniformes: Cyprinidae) larvae and its modulation by seed of Devil's Horsewhip *Achyranthes aspera*. Acta Ichthyol Piscat 43:119–126. https://doi.org/10.3750/AIP2013.43.2.04

Singh A, Sharma J, Paichha M, Chakrabarti R (2019) *Achyranthes aspera* (Prickly chaff flower) leaves-and seeds-supplemented diets regulate growth, innate immunity, and oxidative stress in *Aeromonas hydrophila*-challenged *Labeo rohita*. J Appl Aquac 32:250-267. https://doi.org/10.1080/10454438.2019.1615594

Singh A, Chakrabarti R, Sharma J (2020) Assessment of immunostimulatory characteristics of *Achyranthes aspera* seeds and leaves supplemented diets in *Labeo rohita* fingerlings. Turkish J Fish Aquat Sci 20:795–805. https://doi.org/10.4194/1303-2712-v20_11_03

Siwicki AK (1989) Immunostimulating influence of levamisole on nonspecific immunity in carp (*Cyprinus carpio*). Dev Comp Immunol 13:87–91. https://doi.org/10.1016/0145-305X(89)90021-9

Srivastava PK, Chakrabarti R (2012) Effect of dietary supplementation of seed of *Achyranthes aspera* on the immune system of *Labeo rohita* fry. Isr J Aquac-Bamid 64:1–6. https://doi.org/10.46989/001c.20615

Thomas J, Jerobin J, Seelan TSJ, Thanigaivel S, Vijayakumar S, Mukherjee A, Chandrasekaran N (2013) Studies on pathogenecity of *Aeromonas salmonicida* in catfish *Clarias batrachus* and control measures by neem nanoemulsion. Aquaculture 396:71–75. https://doi.org/10.1016/j.aquaculture.2013.02.024

Umamaheswari M, Dhinesh S, Asokkumar K, Sivashanmugam T, Subhadradevi V, Puliyath J, Madeswaran A (2012) Anticataractic and antioxidant activities of *Abrus precatorius* Linn. against calcium-induced cataractogenesis using goat lenses. Eur J Exp Biol 2:378–384

Varuna KM, Khan M, Sharma PK (2010) Review on *Achyranthes aspera*. J Pharm Res 3:714–717

Zhang A, Chen D, Wei H, Du L, Zhao T, Wang X, Zhou H (2012) Functional characterization of TNF-α in grass carp head kidney leukocytes: induction and involvement in the regulation of NF-κB signaling. Fish Shellfish Immunol 33:1123–1132. https://doi.org/10.1016/j.fsi.2012.08.029

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360335300

Effect of Wind Loads on Low-Rise Pitched and Circular Arch Roof Structures: A Comparative Study based on Numerical Simulation

Article *in* Civil Engineering and Architecture · May 2022

DOI: 10.13189/cea.2022.100329

CITATIONS
0

READS
89

3 authors:



Aditya Kumar Jha
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE



Amartya Sinha
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Ritu Raj Ahirwar
Delhi Technological University
31 PUBLICATIONS **78** CITATIONS
SEE PROFILE

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

Wind effects on tall buildings View project

All content following this page was uploaded by Ritu Raj Ahirwar on 03 May 2022.

The user has requested enhancement of the downloaded file.

Civil Engineering and Architecture 10(3): 1129-1141, 2022 http://www.hrpub.org
DOI: 10.13189/cea.2022.100329

Effect of Wind Loads on Low-Rise Pitched and Circular Arch Roof Structures: A Comparative Study based on Numerical Simulation

Aditya Kumar Jha, Amartya Sinha, Ritu Raj*

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

Received February 3, 2022; Revised March 4, 2022; Accepted April 15, 2022

Cite This Paper in the following Citation Styles

(a): *[1] Aditya Kumar Jha, Amartya Sinha, Ritu Raj, "Effect of Wind Loads on Low-Rise Pitched and Circular Arch Roof Structures: A Comparative Study based on Numerical Simulation," Civil Engineering and Architecture, Vol. 10, No. 3, pp. 1129-1141, 2022. DOI: 10.13189/cea.2022.100329.*

(b): *Aditya Kumar Jha, Amartya Sinha, Ritu Raj (2022). Effect of Wind Loads on Low-Rise Pitched and Circular Arch Roof Structures: A Comparative Study based on Numerical Simulation. Civil Engineering and Architecture, 10(3), 1129-1141. DOI: 10.13189/cea.2022.100329.*

Copyright©2022 by authors, all rights reserved. Authors agree that this article remains permanently open access under the terms of the Creative Commons Attribution License 4.0 International License

Abstract Through numerical simulations based on Computational Fluid Dynamics (CFD), this study examines the effect of wind load operating on a pitched roof and a circular arch roof of an identical plan area of a storage structure with one wall opening. Many numerical simulations have been performed by other researchers for other types of roofs to understand the behavior of wind flow associated with them. Also, the wind standards of multiple countries have provided the Coefficients of Pressures for standard building shapes. However, there is lack of similar work for structures having one wall opening. Ansys CFX fluid flow software has been used to carry out the simulations using a standard k-ε turbulence model. The simulations have been carried out for 7 wind incidence angles at an interval of 30°. Pressure, Force, Moment, Drag and Lift Coefficients generated on the roofs were compared. It was found that, although variation in these coefficients w.r.t the wind incidence angles are similar for both the roofs types, the sharp-edged rooftop of the pitched roof contributes to higher magnitudes of these coefficients as compared to the smooth arch roof. Mathematical modules between the face average of pressure coefficients and the wind incidence angles were established in a polynomial form which could be used to find the face average of pressure coefficients and corresponding pressures and forces for arbitrary wind incidence angles.

Keywords Low Rise Structures, Pitched Roof, Arch Roof, Ansys CFX, Wind Load, Pressure Coefficients

1. Introduction

The global economic expansion and large-scale industrialization have led to rising in the requirement of structures for storage facilities such as warehouses, godowns, cold storage, bunkers, structures like motor workshops, manufacturing units, etc. The air transportation network also plays an important role in facilitating rapid globalization. Hangars are key structures at airports that are used for aircraft repair, storage, and upkeep. These structures have large plan dimensions and involve truss roofs. The influence of wind on such roofs is enormous and needs to be critically analyzed. In recent years, various failures of such roof structures due to strong winds have occurred. In 2018, at Hobby airport in Huston, strong winds damaged the hangar roof damaging the parked planes.

The geometry of the roof and the slope of the roof are both important elements in the structure's wind load resistance. As a result, predicting wind loads on roofs is essential for building safe and cost-effective roof structures and trusses. With the use of computational fluid dynamics (CFD) simulations, wind load estimations may be done using numerical solutions for wind pressures in which a

model of the structure is subjected to a wind flow profile in an artificially built wind tunnel domain. The current research was conducted as a first step toward using numerical simulation to examine the wind influence on the roofs of a model structure. The time averaged Reynold's Average Navier Stokes (RANS) equations of continuity and momentum were investigated in this work by using the Ansys CFX solver with the usual k – ε turbulence model. It is the most basic type of numerical simulation, requiring boundary conditions similar to natural conditions. RANS differential equations are solved along with discretized domain and bluff body elements, and the procedure is repeated several times. The flow parameter results may be viewed, and post-result data can be collected. The same process can be done in wind tunnel experiments also, but it is time consuming and the cost of experiment also increases.

Since the effects of wind loading on a structure are unique and wind flow behaves according to the structure's profile, the wind loads show varying patterns on different localized areas on the faces and roof of buildings. Various studies, both analytical, as well as numerical, have been done on different types of structures. Blackmore [1], Gerhardt & Kramer [2], Meecham [3] and Mehta et al. [4] have carried out Wind Flow measurements on actual low-rise buildings of different shapes, and side ratios and have found that small modifications in the edges and shapes of the walls largely impact the wind flow.

Amin and Ahuja[5], Paul and Dalui [6], Mallick et al. [7], Bhattacharyya et al. [8], Chakraborty et al. [9], Mukherjee et al. [10], Bairagi and Mukherjee [11], Sanyal and Dalui [12,13], Kumar and Raj [14] and Gaur and Raj [15] have worked on high rise buildings of various shapes like L, T, Z, C, E, +, N, Y, Octagonal-Oval, etc. by using experimental as well as numerical studies. It has been found that different plan shapes result in altogether different wind pressure patterns.

Singh and Roy [16-19] have worked on CFD simulation on wind effects over Pyramidal Roofs and have studied the effect of wind direction and roof slope for single and double storied pyramidal roof buildings. Rani and Ahuja [20,21], Roy et al. [22], Sakib et al. [23] have studied Canopy roofs subjected to wind loads by using Wind Tunnel Testing. Verma and Ahuja [24] and Zhou, et al. [25] have studied wind effects on Dome structures, which provided a unique shape of the roofs, using wind tunnel tests CFD analysis respectively. Verma and Ahuja [26] have studied the effect of cylindrical roofs in single span as well as double span configurations.

Ahmad and Kumar [27] have studied interference effects on wind load for hip roofs by conducting wind tunnel tests. Significant enhancements and shielding effects were observed for different configurations. Stathopoulos [28] and Stathopoulos et al. [29] have studied the effect of a tall nearby building on the wind loading of low buildings and also carried out wind tunnel tests to examine the effect of the location of trees in the vicinity of a low-rise building. Significant changes in wind pressures were observed. Tecle et al. [30] have studied the Opening and Compartmentalization Effects of Internal Pressure in Low-Rise Buildings with Gable and Hip Roofs. Gaur et al. [31] have worked on the interference effect on corner configured structures using CFD for different geometry and blockage configurations.

In this paper, a comparative study between Pitched Roof and Arch Roof over a low-rise rectangular plan building was performed using CFD simulations performed under the standard k-ε turbulence model for high turbulence intensity in the Ansys CFX fluid flow package. It is a preliminary examination of wide span low-rise structures for determining the suitability of use of Ansys CFX. For this study, an arbitrary dimension of the structure has been taken. The calculations of pressure, force, moment, drag and lift coefficients for the models were used for comparison and polynomial mathematical modules were developed between the average pressure coefficient (C_p) values and the wind incidence angles were established which can be used to find average C_p values and corresponding pressures and forces for arbitrary wind incidence angles.

2. Simulation Model and Discretization of Computational Domain

For the analysis, a structure with a 1 m x 0.6 m plan area, and 0.5 m wall height is considered as shown in Figures 1 and 2. The pitched roof is inclined at an angle of 20° with the horizontal which results in a maximum elevation of 0.6092 m at the center of the roof. For comparison with an arch roof, a circular arch is considered as a roof that has the same central elevation. The models are created with a wall thickness of 0.01 m and one side of the models is kept open, in order to subject the structure to internal wind pressure also. A scale of 1:10 has been adopted in this study for the scope of future experimental verification.

Equations (1) and (2) represent the profile of the pitched roof and circular arch roof:

$$z = 0.6092 - 0.364|x| \quad (1)$$

$$z = 0.1425 + \sqrt{0.4667^2 - x^2} \quad (2)$$

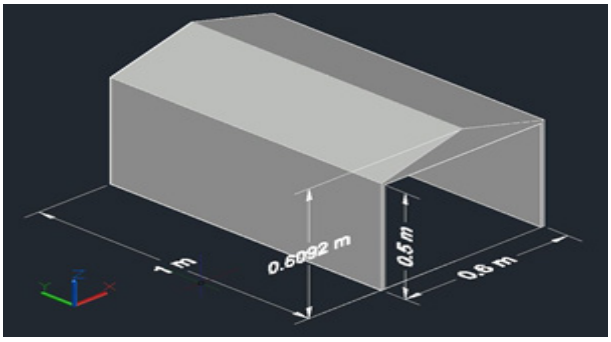


Figure 1. Pitched Roof Model

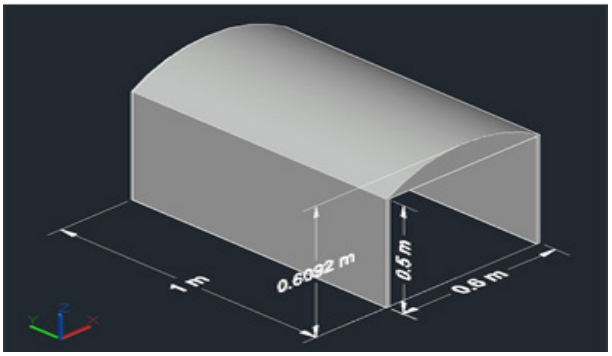


Figure 2. Circular Arch Roof Model

The process of simulation in Ansys CFX requires the creation of discretized flow domain around the geometry and involves the setup for natural boundary conditions which are solved through numerical simulations to get the results. Before the actual simulation, the flow setup and the discretized volume mesh need to be validated against standard wind code provisions. For validating the flow setup, a validation model without opening from one side of the same plan area and height as shown in Figure 3 was subjected to a uniform homogeneous wind flow of 10 m/s at the inlet of the domain since the considered structure is low-rise and the wind flow around the structure is the part of Atmospheric Surface Layer, where the velocity profile can be treated as a uniform profile for simplicity in the analysis. In the validation as well as simulation models, Roof Surface A and B are the windward and leeward sides of the roof as shown in Figure 4.

After a number of trials, the size for the domain volume mesh grid was finalized as 0.2 m and the model faces and roof meshing sizes as 0.05 m; much smaller as the focus in the present study is to map the pressure effect on the roofs of the model. A diagrammatical representation of the grid size variation is shown in Figure 5. Free slip walls were used to define the domain sides and top for which no shear stress is acting at the wall surface but the velocity along the wall surface has a finite value. The component of wind velocity normal to the wall surface, on the other hand, is zero. The ground and the faces of the model were defined as no slip walls imply that the component of wind velocity along the wall's boundary is zero. The outlet was defined as a boundary having 0 Pa relative pressure.

According to Franke et al. [32], the computational domain for the simulation has an optimum size so that the fluid interaction with the model is not harmed by reflection of fluid flow from the domain boundary. For this, the blockage ratio offered by the bluff body should remain below 3%. As a result, the intake and outlet distances from the nearest surface of the model should be kept at 5H and 15H, respectively, and the sides of the domain from the side faces of the model and the top clearance from the model roof should be kept as 5H where H represents the model's maximum height. Using these recommendations, a domain was created for the simulation as shown in Figure 6.

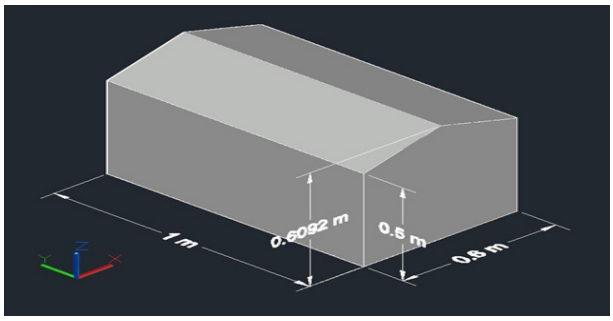


Figure 3. Validation Model

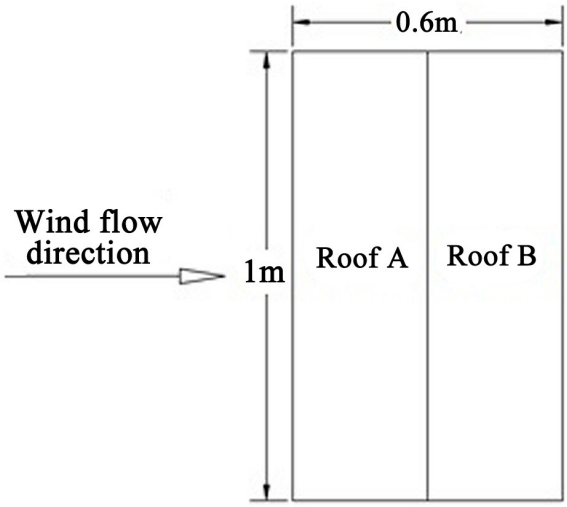


Figure 4. Orientation of Validation Model

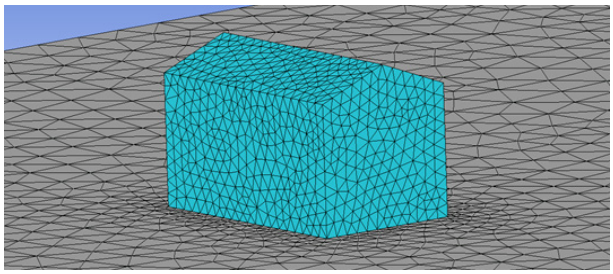


Figure 5. Validation Model - Mesh Diagram

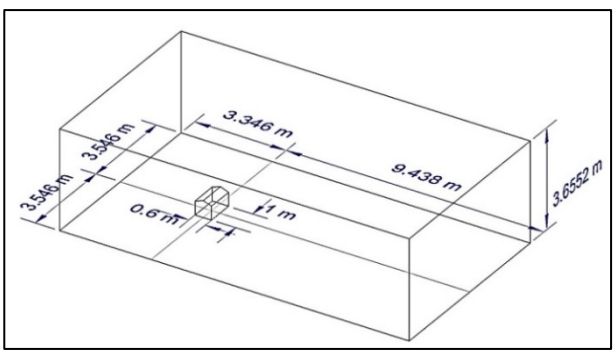


Figure 6. Validation Model - Geometry Setup

The face average value of pressure coefficient obtained

on the validation model was compared with the values given in Indian, British, Australian/ New Zealand, and ASCE standards and the average % variation was found to be 16.5 %, 12.3%, 7.5%, and 7.5% respectively. It can be seen from Figure 7 that the coefficient of pressure values in various international codes/standards deviate from each other and CFD values are within reasonable bounds of 15 to 20%.

For experimental verification, comparison has been made with results available through the wind tunnel tests carried out by Ozmen et al. on pitched roofs to study the effect of roof pitch on the mean pressure coefficient along the mid-axis of building models [33]. The structure considered for the experimental verification has 100mm x 50mm external dimensions with a 15° roof pitch angle with a wall height of 40 mm and a geometric scale of 1:100. The geometric parameters closely match with that considered in the numerical modeling. The data inferred from the graph of the experimental study shows the mean pressure coefficient (C_{pe}) for the roof faces to be approximately -0.65. This is in close agreement with the numerical simulation result.

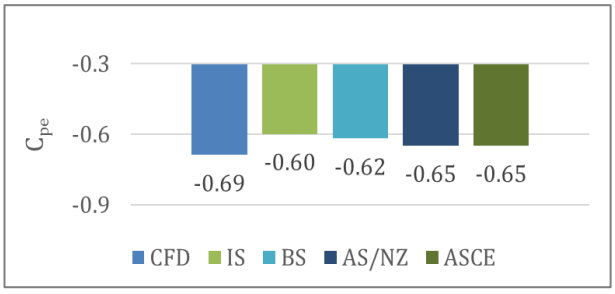


Figure 7. Comparison of C_{pe} obtained with Standard C_{pe} values given by different Wind Codes

Once the flow setup and discretization of the flow domain were validated, the models were subjected to the same wind flow setup with a similar CFD grid. The wind incidence angles were varied from -90° to 0° (clockwise rotation) and +90° to 0° (anticlockwise rotation) at an interval of 30° with 0° direction along the positive x axis which is normal to the symmetric axis of the building as shown in Figure 8.

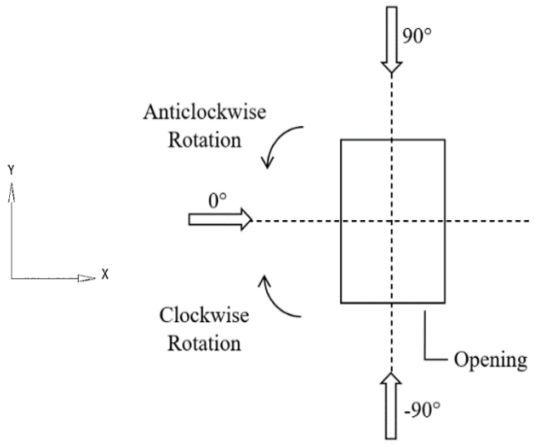


Figure 8. Model Orientation

The further sections deal with the calculations of pressure, force, moment, drag and lift coefficients for the models for comparison.

3. Pressure Coefficients

Pressure Coefficient is a dimensionless quantity which provides a numerical representation for the relative pressure at any point w.r.t the design pressure. It is defined as shown in (3).

$$Pressure\ Coefficient = C_p = \frac{Pressure}{\frac{1}{2}\rho \cdot V^2} \qquad (3)$$

where, ρ & V are the density of air and design wind speed respectively.

The C_p is a point specific value, which is calculated at the vertices of each node in Ansys CFX. If the surface in consideration is external or internal, the corresponding values of pressure coefficients are represented as C_{pe} or C_{pi} respectively and the average pressure coefficient on the surface is represented as $C_{pe(avg)}/C_{pi(avg)}$. The maximum pressure coefficient, $C_{pe(max)}/C_{pi(max)}$, and the minimum pressure coefficient, $C_{pe(min)}/C_{pi(min)}$, at a point on the surface are critical for design of roof elements and supporting structures and fixtures. Based on equation (3), the contours of C_p for the roof surfaces in this study are shown in the following figures.

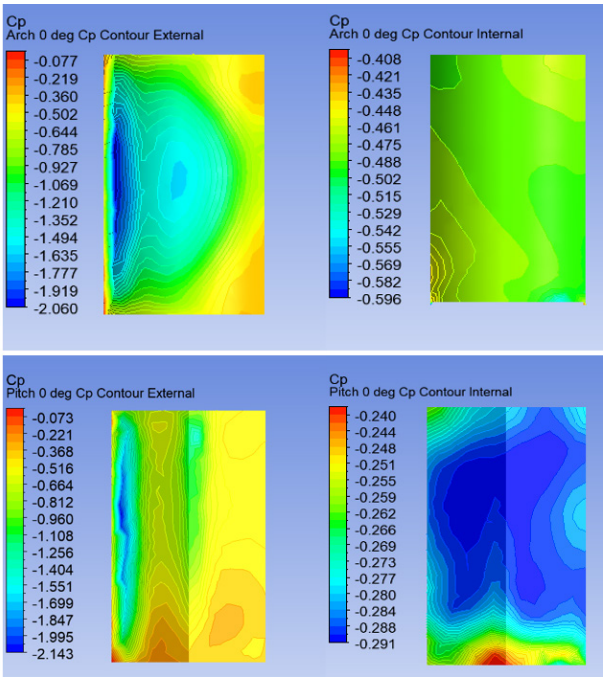


Figure 9. C_p Contours for 0° Wind Incidence

For a 0° wind incidence, Figure 9 depicts the C_p contours for the external and internal faces of the Arch and Pitch roofs. The C_{pe} is negative for both the pitched and arch roofs, with a slight rise in the $C_{pe(min)}$ for the pitched roof. This is due to the sharp edge of the pitch roof's rooftop, which causes eddies on the leeward side of the roof, producing increased suction. In this case, the effect of vortex shedding is easily noticeable. The C_{pi} contour for both roofs has a uniform nature, with the pitched roof having a smaller magnitude of suction C_{pi} values than the arch roof which is due to the presence of a planar roof surface that gives more resistance to air circulation in the internal environment.

The C_p contours in Figures 10, 11, 12 and 13 are obtained corresponding to the wind incidences at +30°, +60° & -30°, -60°. For +30°, higher magnitudes of C_{pe} as well as C_{pi} are observed than the corresponding values for +60°, for both the roofs. This is due to the higher blockage offered by the structures to the wind flow in 30° configuration. The same trend is observed for C_{pe} in -30° and -60° configurations also. However, this trend reverses for C_{pi} in -30° and -60°, since the internal pressures in these cases are governed by the direct wind entry in the internal environment and -60° configuration offers a more exposed opening to the wind flow.

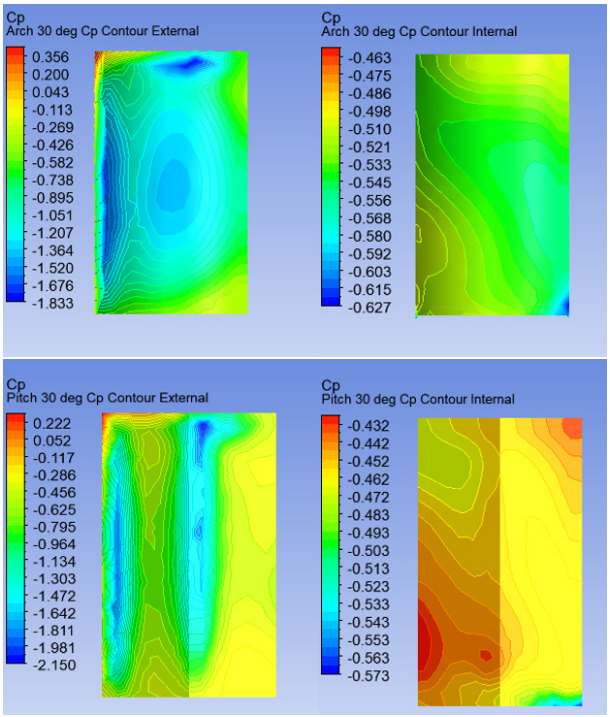


Figure 10. C_p Contours for 30° Wind Incidence

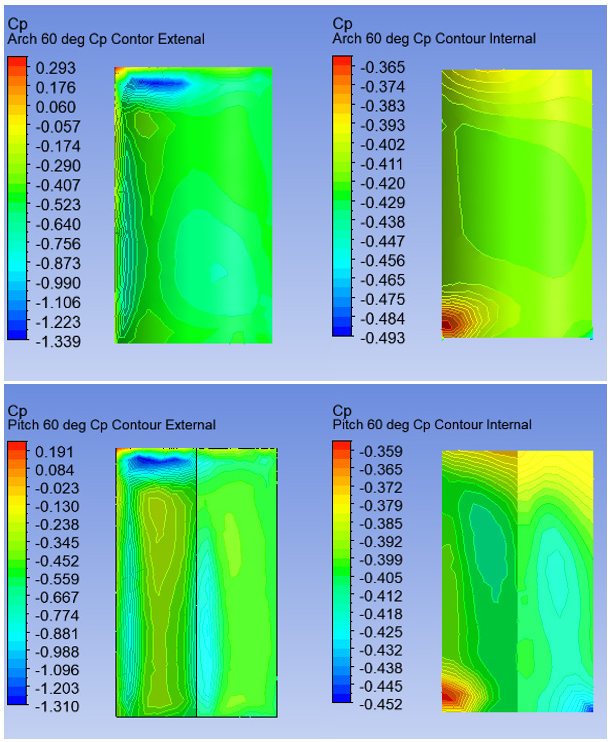


Figure 11. C_p Contours for 60° Wind Incidence

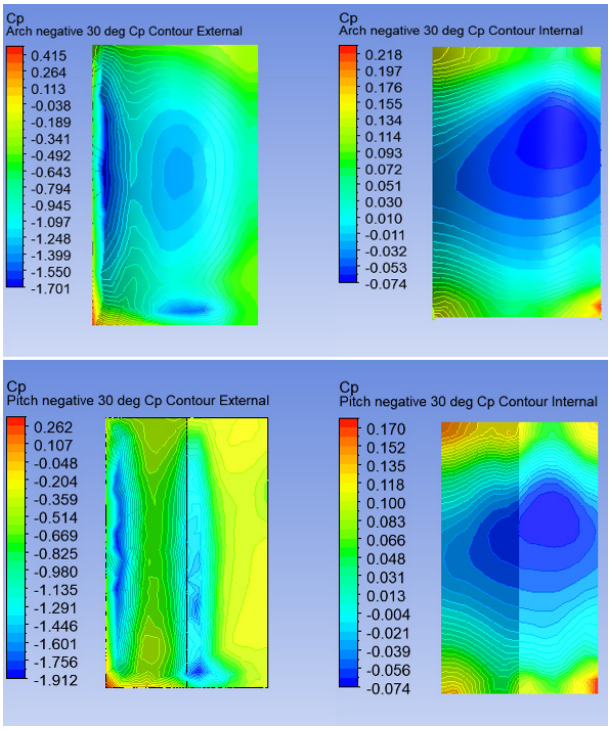


Figure 12. C_p Contours for -30° Wind Incidence

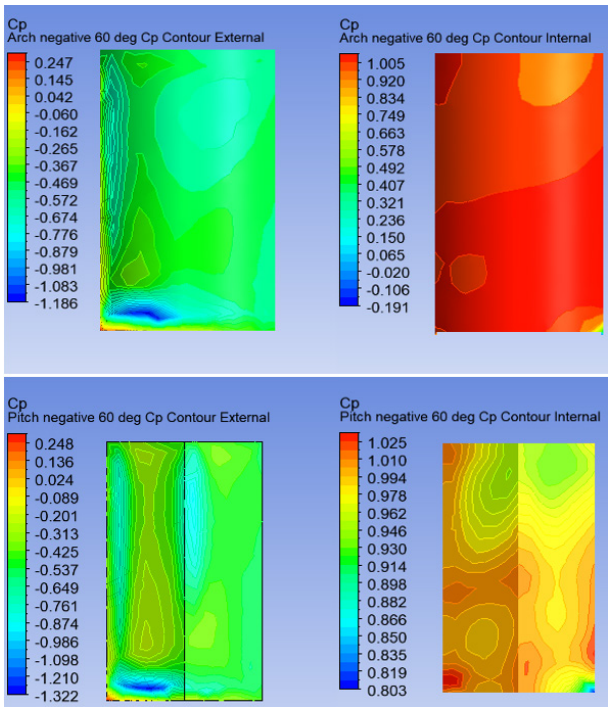


Figure 13. C_p Contours for -60° Wind Incidence

If the arch and pitched roofs are compared among themselves, the C_{pi} does not show significant variation. However, $C_{pe(max)}$ for arch roofs are about 1.75 times that for a pitched roof. The surfaces of arch roofs have more elevation due to their convex upwards nature, which results in more blockage to the wind flow. On the other hand, $C_{pe(min)}$ shows the opposite trend as the suction on the external roof is caused by the flow separation, which occurs more for the pitched roof due to its sharp roof top, thus causing more suction.

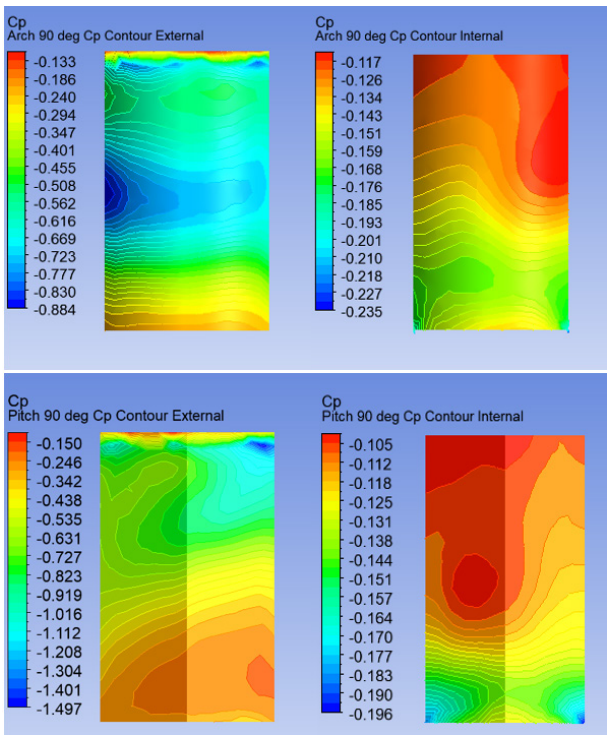


Figure 14. C_p Contours for 90° Wind Incidence

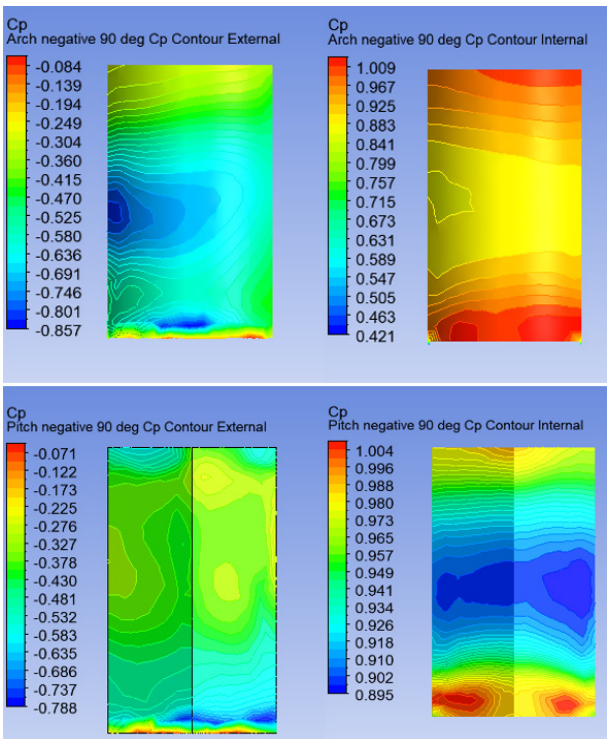


Figure 15. C_p Contours for -90° Wind Incidence

The C_p contours in Figures 14 and 15 are obtained corresponding to the wind incidence for +90° and -90°, respectively. These scenarios depict the building's closed

and open sides with relation to the domain inlets, respectively. For both wind incidences, the contour variation for C_{pe} of both roofs shows negative values throughout the roof surface. Due to the bluff body nature of the pitched roof and the splitting of the wind streamline at the rooftop, the $C_{pe(min)}$ value for the external roof faces for both wind incidence angles are higher for the pitched roof.

The values of C_{pi} are having smaller magnitudes and are causing suction for +90° wind incidence since the wind circulates in the internal environment after losing a significant amount of velocity due to the presence of a closed wall on the windward side. However, the same is positive for -90° because of the outward pressure exerted on the internal roof face due to the opening in the structure being on the windward side. The $C_{pi(max)}$ values for both pitched and arch roof internal surfaces are identical (1.009 and 1.004 respectively), but the $C_{pi(min)}$ value for the -90° wind incidence is nearly twice in the pitched roof case due to the pitched roof's sharp edge and planar surface.

The maximum, minimum and average C_p were plotted for the External and Internal Roof Surfaces for both the structures for different wind incidence angles to study the variation of pressure coefficients for each roof surface as shown in Figures 16 to 21.

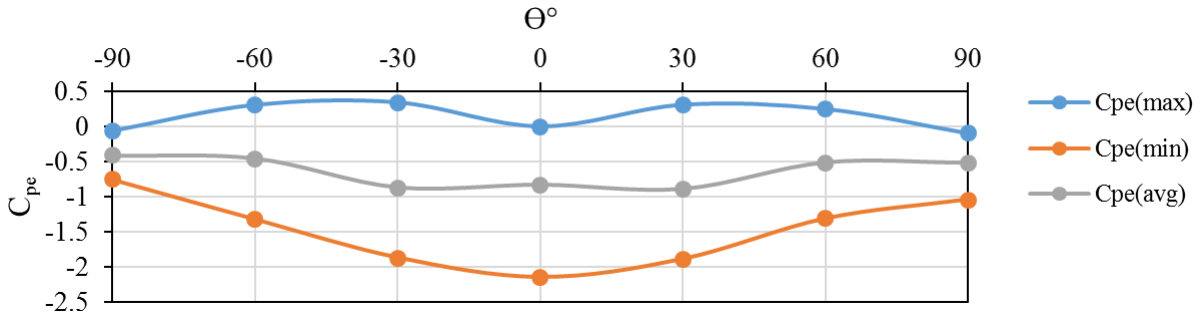


Figure 16. Pitched Roof A - External Coefficient of Pressure (C_{pe}) v/s Wind Incidence Angle (Θ°)

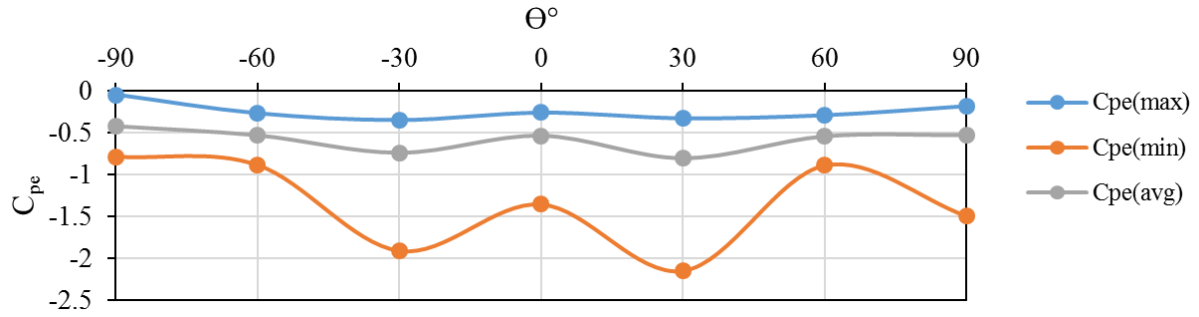


Figure 17. Pitched Roof B - External Coefficient of Pressure (C_{pe}) v/s Wind Incidence Angle (Θ°)

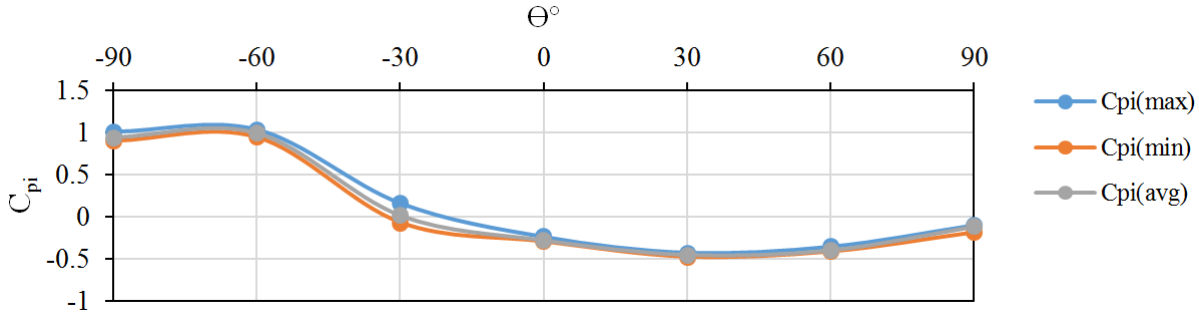


Figure 18. Pitched Roof A - Internal Coefficient of Pressure (C_{pi}) v/s Wind Incidence Angle (Θ°)

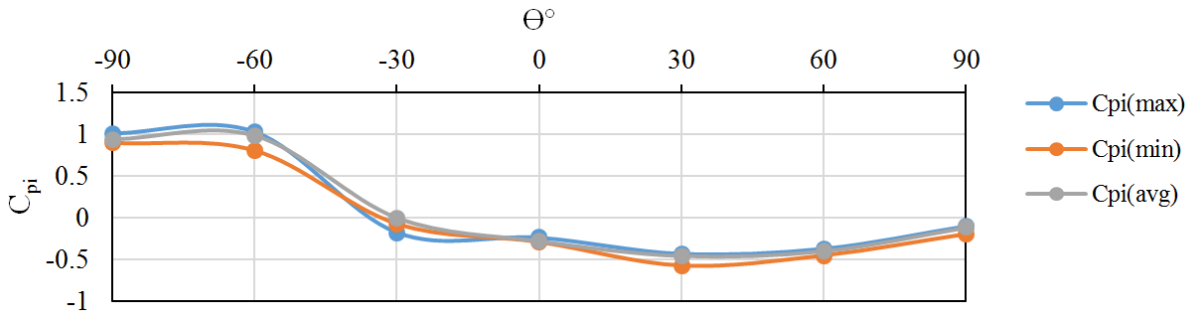


Figure 19. Pitched Roof B - Internal Coefficient of Pressure (C_{pi}) v/s Wind Incidence Angle (Θ°)

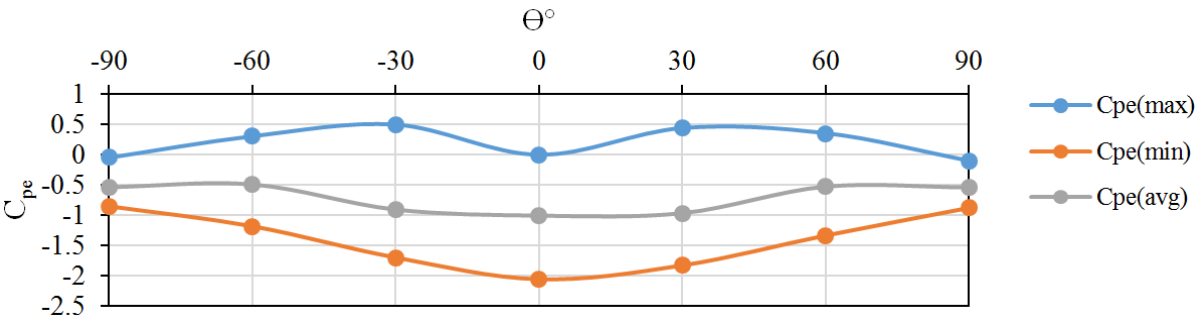


Figure 20. Arch Roof - External Coefficient of Pressure (C_{pe}) v/s Wind Incidence Angle (Θ°)

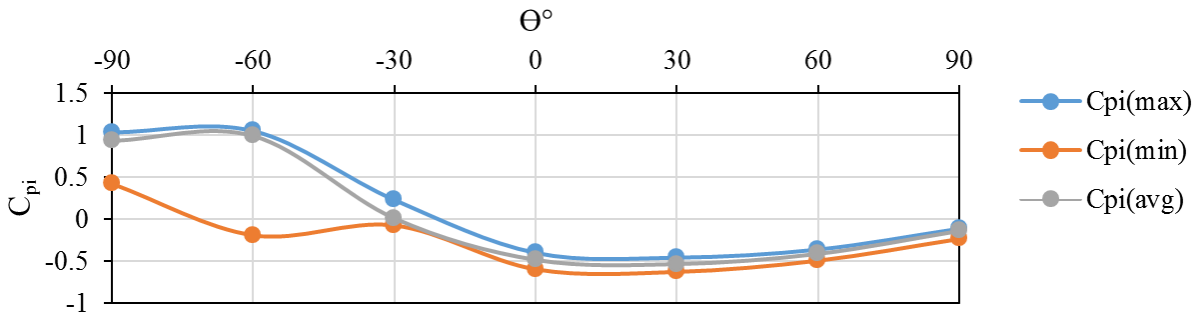


Figure 21. Arch Roof - Internal Coefficient of Pressure (C_{pi}) v/s Wind Incidence Angle (Θ°)

There is a symmetry about 0° wind incidence angle for the $C_{p(avg)}$, $C_{p(max)}$, and $C_{p(min)}$ graphs of external faces for both the roof types with the value of $C_{p(avg)}$ increasing for the pitched roof by nearly 1.2 times at 0° wind incidence because of its sharp edge. The nature of the C_p graphs for the internal faces is also similar for both the roofs. However, there is no symmetry about the 0° wind incidence angle as the effect of closed and open walls becomes more eminent for the internal roof faces. The average C_p values of the internal faces for both roofs are almost identical in this case.

4. Force, Moment, Drag and Lift Coefficients

Force Coefficients (C_{fx}, C_{fy}. C_{fz}) as shown in (4), (5), and (6) and Moment Coefficients (C_{mx}, C_{my}, C_{mz}) as shown in (7), (8) and (9) are important parameters for the design of buildings and roofs against wind loadings. These coefficients provide a platform to compare the behavior of the structures against wind. They can be applied to prototype building for calculating resultant forces and moments acting on the buildings in different directions. The drag and lift coefficients (C_d and C_l) as shown in (10) and (11) are the force coefficients in the direction along and perpendicular to the direction of wind flow. These coefficients were calculated for the different incidence angles of wind for the study of the two roofs.

$$C_{fx} = \frac{F_x}{\frac{1}{2}.\rho.V^2.L_y.H} \quad (4)$$

$$C_{fy} = \frac{F_y}{\frac{1}{2}.\rho.V^2.L_x.H} \quad (5)$$

$$C_{fz} = \frac{F_z}{\frac{1}{2}.\rho.V^2.L_x.L_y} \quad (6)$$

$$C_{mx} = \frac{M_x}{\frac{1}{2}.\rho.V^2.L_x.H^2} \quad (7)$$

$$C_{my} = \frac{M_x}{\frac{1}{2}.\rho.V^2.L_y.H^2} \quad (8)$$

$$C_{mz} = \frac{M_x}{\frac{1}{2}.\rho.V^2.L_x.L_y.H} \quad (9)$$

$$C_d = \frac{F_{along}}{\frac{1}{2}.\rho.V^2.L_y.H} \quad (10)$$

$$C_l = \frac{F_{across}}{\frac{1}{2}.\rho.V^2.L_x.H} \quad (11)$$

where, ρ & V are the density of air and wind velocity respectively; L_x is the projected length along wind direction and L_y is the projected length across wind direction which depend on the orientation of the model as described in Table 1; H is the maximum height of the model (=0.6092 m); F_x, F_y, F_z & M_x, M_y, M_z are the sum of the forces and moments induced on all the faces of the building along the x, y and z directions; F_{along} and F_{across} are the forces on the building along the wind direction and perpendicular (across) wind direction respectively.

Table 1. Projected Lengths of the buildings about x and y directions

Angle	L_x (m)	L_y (m)
0°	0.6	1
±30°	1.0196	1.1660
±60°	1.1660	1.0196
±90°	1	0.6

Figures 22 and 23 represent the variation of coefficients of forces along the different directions w.r.t the wind incidence angles for the two buildings. The nature of the graphs is almost similar for both the roof types and symmetry about the 0° wind incidence angle can also be seen for the C_{fx} graph. However, the values of C_{fx} and C_{fy} are slightly lower (0.133 and 0.78 times respectively) at 0° wind incidence angle for the arch roof building as compared to the pitched roof because of its plane and continuous roof top.

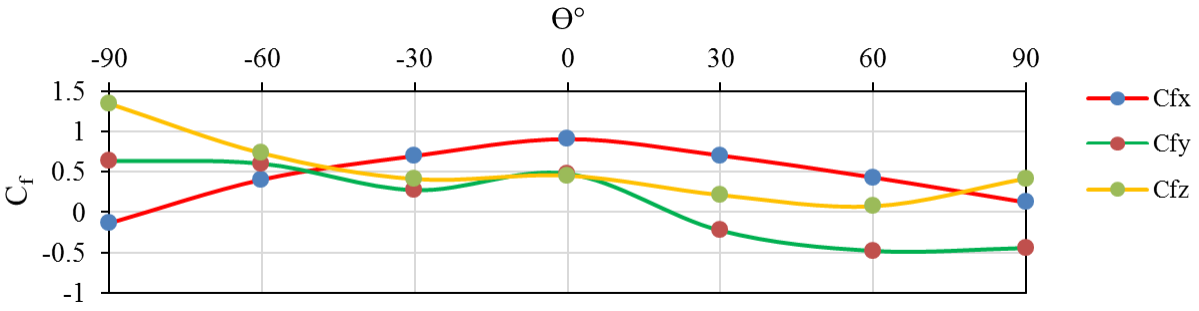


Figure 22. Pitched Roof - Force Coefficient (C_f) v/s Wind Incidence Angle (Θ°)

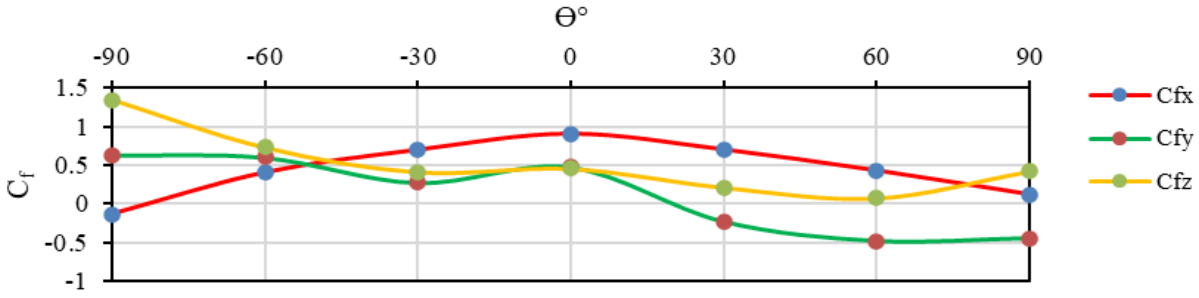


Figure 23. Arch Roof - Force Coefficient (C_f) v/s Wind Incidence Angle (Θ°)

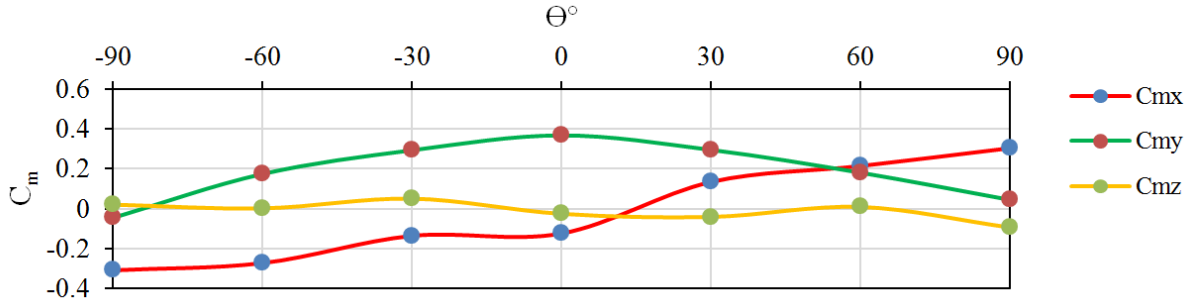


Figure 24. Pitched Roof - Moment Coefficient (C_m) v/s Wind Incidence Angle (Θ°)

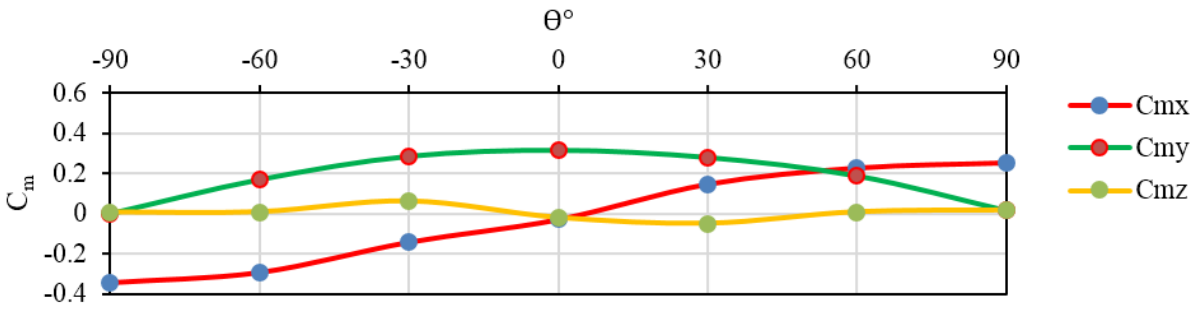


Figure 25. Arch Roof - Moment Coefficient (C_m) v/s Wind Incidence Angle (Θ°)

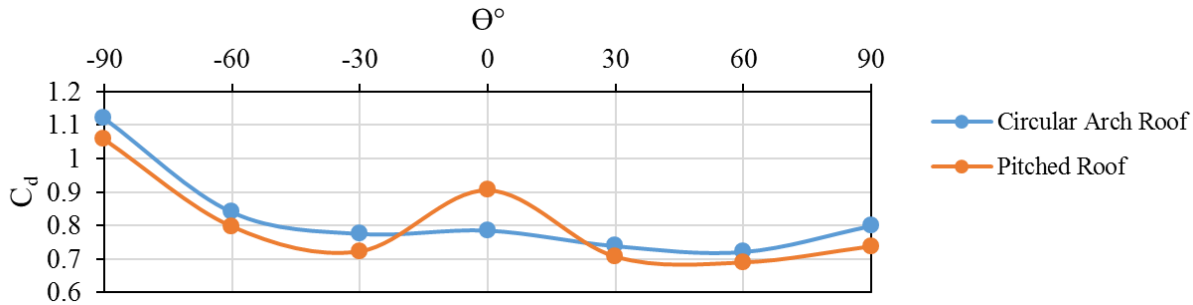


Figure 26. Drag Coefficient (C_d) v/s Wind Incidence Angle (Θ°)

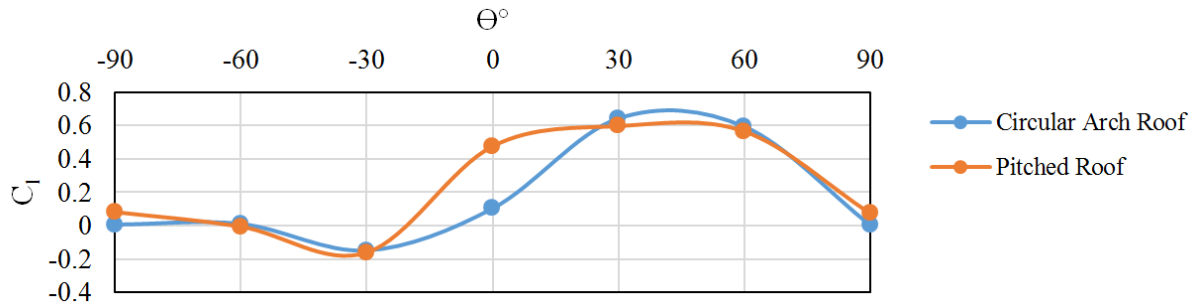


Figure 27. Lift Coefficient (C_l) v/s Wind Incidence Angle (Θ°)

Figures 24 and 25 represent the variation of coefficients of moments along the different directions w.r.t the wind incidence angles for the two buildings. The nature of the graphs is similar for both the roof types and symmetry about the 0° wind incidence angle can also be seen for each of the C_{mx}, C_{my} and C_{mz} graphs. The magnitude of the data points is also somewhat similar for the C_{my} and C_{mz} graphs however a decrease in the C_{mx} value by nearly 4 times can be seen at the 0° wind incidence angle for the pitch roof building as compared to the arch roof. It is also observed that there exists a resemblance in the C_{fx}, C_{my} graph and C_{fy}, C_{mx} graph respectively in terms of their nature. This occurs because of the similar variables in their equations.

Figures 26 and 27 represent the coefficient of drag and lift for the two roof types w.r.t the wind incidence angles. The nature of both C_d and C_l graphs is almost similar with very slight differences in magnitudes at the different wind incidence angles for both the roof types, however a significant increase in the values of both C_d and C_l is seen at 0° for the pitch roof building (by nearly 1.15 and 4.60 times respectively) and this happens because of its sharp roof edge which induces a larger force in the along wind direction (drag) and splits the wind streamline for a larger wind force in the across direction (lift).

5. Mathematical modules between $C_{pe(avg)}/C_{pi(avg)}$ and Wind Incidence Angle

The $C_{pe(avg)}/C_{pi(avg)}$ face average values of pressure coefficients are important to study the behavior of the roof under wind load as other parameters such as pressures and forces could be easily estimated once these values are obtained. However, the process to obtain these values for different wind incidence angles on Ansys CFX is a tedious and cumbersome process, since the software reiterates from the beginning. Thus, a standard relationship established between the $C_{pe(avg)}/C_{pi(avg)}$ values and the wind incidence angles is of great use. Here the available values of $C_{pe(avg)}/C_{pi(avg)}$ for the seven wind incidence angles

through the simulations conducted are utilized to establish polynomial relations. These polynomial relations are developed by studying the nature of the $C_{pe(avg)}/C_{pi(avg)}$ graphs for both the external and internal faces of both the roofs and are given in Table 2. The coefficients are solved through the use of the available boundary conditions ($C_{pe(avg)}/C_{pi(avg)}$ values of different wind incidence angles) and verified through their trendline plots. Suitable periods of angles are classified so as to satisfy the nature of the curve variation.

6. Conclusion and Future Work

A comparison of a pitched roof and a circular arch roof low-rise structure has been made in this study. Both types of roofs are common in the industry, and they feature similar frontal projections. When the building's dimensions and slope of the roofs are kept constant, the shape and curvature of the roofs govern the wind load response. The opening on one side of the building causes the external and interior roof faces to behave differently.

While the sharp edges of the pitch roof produce slight increase in suction on external roof for the 0°, it is observed that the clockwise rotation of the building with respect to the wind causes an increase in C_{pi} values for both roofs, corresponding to larger exposed opening to the wind and with higher magnitudes for the pitched roof due to its sharp edge. The anticlockwise rotations, on the other hand, produce negative C_{pe} and C_{pi} values, with the arch roof having a larger magnitude than the pitched roof.

The force and moment coefficients computed along the x, y and z directions show similar nature for both roofs with symmetries about the 0° wind incidence for these coefficients. However, a lower value of C_{fx} and C_{fy} is observed for the arch roof than in pitched roof. This is because the velocity sharply escapes the sharp roof edge and vortex shedding becomes more eminent in the pitched roof which leads to a relatively higher magnitude of suction force in the pitched roof, which increases the forces in the horizontal plane.

The drag coefficient (C_d) and lift coefficient (C_l) show similar variation for both roofs, but the magnitudes are significantly higher for the pitched roof, especially at 0° wind incidence angle. This roof's sharp-edged roof top is again responsible for the increase in coefficients due to better wind streamline splitting.

The relations established between the $C_{pe(avg)}/C_{pi(avg)}$ values and the wind incidence angles in polynomial form can be of great use for quicker estimation of average pressure coefficient, pressures, and forces on both the external and internal faces of both the roofs.

Table 2. Equations for $C_{pe(avg)}/C_{pi(avg)}$ in terms of θ (in radians)

Roof	Equation for $C_{pe(avg)}/C_{pi(avg)}$ in terms of θ (in radians)	Range for θ
Arch Roof External Face	$C_{pe(avg)} = -0.027\theta - 0.5040$	θ ∈ [-90°, -60°] & [60°,90°]
	$C_{pe(avg)} = 0.0482\theta^3 + 0.4816\theta^2 - 0.0702\theta - 1.0113$	θ ∈ [-60°,60°]
Arch Roof Internal Face	$C_{pi(avg)} = -0.110\theta - 1.1040$	θ ∈ [-90°, -60°]
	$C_{pi(avg)} = -0.1606\theta^3 + 0.6938\theta^2 - 0.4874\theta - 0.483$	θ ∈ [-60°,90°]
Pitched Roof External Face A	$C_{pe(avg)} = -0.6733\theta^2 - 1.8423\theta - 1.6497$	θ ∈ [-90°, -30°]
	$C_{pe(avg)} = -0.1775\theta^2 - 0.0201\theta - 0.8316$	θ ∈ [-30°,30°]
	$C_{pe(avg)} = -0.6952\theta^2 + 1.8105\theta - 1.6482$	θ ∈ [-30°,90°]
Pitched Roof Internal Face A	$C_{pi(avg)} = 0.096\theta + 1.096$	θ ∈ [-90°, -60°]
	$C_{pi(avg)} = 0.1544\theta^4 - 0.288\theta^3 + 0.376\theta^2 - 0.3539\theta - 0.2802$	θ ∈ [-60°,90°]
Pitched Roof External Face B	$C_{pe(avg)} = -0.1813\theta^2 - 0.683\theta - 1.0484$	θ ∈ [-90°, -30°]
	$C_{pe(avg)} = -0.8594\theta^2 - 0.0619\theta - 0.5373$	θ ∈ [-30°,30°]
	$C_{pe(avg)} = -0.4497\theta^2 + 1.2049\theta - 1.3129$	θ ∈ [30°,90°]
Pitched Roof Internal Face B	$C_{pi(avg)} = 0.088\theta + 1.073$	θ ∈ [-90°, -60°]
	$C_{pi(avg)} = 0.1646\theta^4 - 0.3016\theta^3 + 0.3636\theta^2 - 0.3371\theta - 0.2833$	θ ∈ [-60°,90°]

The results and conclusions established in this study have direct implications on similar studies on the other roof types and varied width to span ratios. This comparative study provides key insights on the difference created by the shape and curvature of the roof on the pressure, force, moment, drag and lift coefficients which can be extended to other families of roofs. This will be beneficial in the study of the structural soundness of the roof types under wind load conditions, accompanied by an economic design. This study can be extended towards high-speed wind conditions also. The relationship developed between the $C_{pe(avg)}/C_{pi(avg)}$ values and wind incidence angle can also be extended to develop relationships between the wind incidence angles and the other coefficients (force, moment, drag, lift) which can help in deciding the best suited roof among the different families of roofs.

Acknowledgments

The authors would like to thank the Department of Civil Engineering, Delhi Technological University for providing the required training and research support for conducting the simulations on Ansys Software.

REFERENCES

[1] P. A. Blackmore. Load reduction on flat roofs - the effect of edge profile, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 29, No. 1, pp. 89–98, 1988.

[2] H. J. Gerhardt and C. Kramer. Effect of building geometry on roof wind loading, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 43, No. 1, pp. 1765–1773, 1992.

[3] D. Meecham. The improved performance of hip roofs in extreme winds — A case study, Journal of Wind Engineering and Industrial Aerodynamics, Vol. 43, No. 1, pp. 1717–1726, 1992.

[4] K. C. Mehta, M. L. Levitan, R. E. Iverson, and J. R. McDonald. Roof corner pressures measured in the field on a low building, Journal of Wind Engineering and Industrial Aerodynamics, pp. 181–192, 1992.

[5] J. A. Amin and A. K. Ahuja. Experimental study of wind-induced pressures on buildings of various geometries, 2011.

[6] R. Paul and S. K. Dalui. Wind effects on 'Z' plan-shaped tall building: a case study, International Journal of Advanced Structural Engineering, vol. 8, no. 3, pp. 319–335, 2016.

[7] M. Mallick, A. Kumar, and K. C. Patra. Experimental Investigation on the Wind-Induced Pressures on C-Shaped Buildings, KSCE Journal of Civil Engineering, Vol. 23, No. 8, pp. 3535–3546, 2019.

[8] B. Bhattacharyya, S. K. Dalui, and A. K. Ahuja. Wind Induced Pressure on 'E' Plan Shaped Tall Buildings, Jordan Journal of Civil Engineering, Vol. 8, No. 2, p. 120, 2014.

[9] S. Chakraborty, K. Dalui, and A. K. Ahuja. Experimental Investigation of Surface Pressure on '+' Plan Shape Tall Building, Jordan Journal of Civil Engineering, Vol. 8, No. 3, p. 251, 2014.

[10] S. Mukherjee, S. Chakraborty, S. K. Dalui, and A. K. Ahuja. Wind induced pressure on 'Y' plan shape tall building, Wind and Structures, An International Journal, Vol. 19, No. 5, pp. 523–540, 2014.

[11] K. Bairagi and A. Mukherjee. Wind pressure and velocity pattern around 'N' plan shape tall building - A case study, Asian Journal of Civil Engineering, Vol. 18, No. 8, pp. 1241–1258, 2017.

[12] P. Sanyal and S. K. Dalui. Effects of internal angle between limbs of 'Y' plan shaped tall building under wind load, Journal of Building Engineering, Vol. 33, 2021.

[13] P. Sanyal and S. K. Dalui. Effects of side ratio for 'Y' plan shaped tall building under wind load, Building Simulation, Vol. 14, No. 4, pp. 1221–1236, 2021.

[14] Kumar and R. Raj. Study of Pressure Distribution on an Irregular Octagonal Plan Oval-Shape Building Using CFD, Civil Engineering Journal, Vol. 7, No. 10, 2021.

[15] N. Gaur and R. Raj. Aerodynamic mitigation by corner modification on square model under wind loads employing CFD and wind tunnel, Ain Shams Engineering Journal, 2021.

[16] J. Singh and A. K. Roy. Effects of roof slope and wind direction on wind pressure distribution on the roof of a square plan pyramidal low-rise building using CFD simulation, International Journal of Advanced Structural Engineering, Vol. 11, No. 2, pp. 231–254, 2019.

[17] J. Singh and A. K. Roy. CFD simulation of the wind field around pyramidal roofed single-story buildings, SN Applied Sciences, Vol. 1, No. 11, 2019.

[18] J. Singh and A. K. Roy. Wind Pressure Coefficients on Pyramidal Roof of Square Plan Low Rise Double Storey Building, Computational Engineering and Physical Modeling, Vol. 2, No. 1, pp. 1–16, 2019.

[19] J. Singh and A. K. Roy. Wind loads on roof of low-rise buildings, Disaster Advances, Vol. 14, No. 5, pp. 83–98, 2021.

[20] N. Rani and A. K. Ahuja. Wind Loads on Multi-Span Mono-Slope Canopy Roof, in Urbanization Challenges in Emerging Economies, pp. 702–708, 2017.

[21] N. Rani and A. K. Ahuja. Wind Loads on Flat Canopy Roofs, 2017.

[22] K. Roy, A. K. Ahuja, and V. K. Gupta. Variation of Wind Pressure on Canopy-Roofs, International Journal of Earth Sciences and Engineering, Vol. 3, No. 2, pp. 19–30, 2010.

[23] F. A. Sakib, T. Stathopoulos, and A. K. Bhowmick. A review of wind loads on canopies attached to walls of low-rise buildings, Engineering Structures, Vol. 230, 111656, 2021.

[24] Verma and A. K. Ahuja. Wind Pressure Distribution on Rectangular Plan Buildings with Multiple Domes, International Journal of Engineering and Technical Research (IJETR), Vol. 3, No. 7, pp. 129–133, 2015.

[25] P. Zhou, B. Tang, P. Liu, and G. Zheng. Study on Wind Load Distribution on the Surface of Dome Structure Based on CFD Numerical Simulation, in Journal of Physics: Conference Series, Vol. 1176, No. 5, 2019.

[26] Verma and A. K. Ahuja. Wind Pressure Distribution on Low-Rise buildings with Cylindrical Roofs, in Implementing Innovative Ideas in Structural Engineering and Project Management, pp. 1211–1216, 2015.

[27] S. Ahmad and K. Kumar. Interference effects on wind loads on low-rise hip roof buildings, Engineering Structures, Vol. 23, No. 12, pp. 1577–1589, 2001.

[28] T. Stathopoulos. Adverse Wind Loads on Low Buildings Due to Buffeting, Journal of Structural Engineering, Vol. 110, No. 10, pp. 2374–2392, 1984.

[29] T. Stathopoulos, D. Chiovitti, and L. Dodaro. Wind shielding effects of trees on low buildings, Building and Environment, Vol. 29, No. 2, pp. 141–150, 1994.

[30] S. Tecle, G. T. Bitsuamlak, and A. G. Chowdhury. Opening and Compartmentalization Effects of Internal Pressure in Low-Rise Buildings with Gable and Hip Roofs, Journal of Architectural Engineering, Vol. 21, No. 1, 04014002, 2015.

[31] N. Gaur, R. Raj, and P. K. Goyal. Interference effect on corner-configured structures with variable geometry and blockage configurations under wind loads using CFD, Asian Journal of Civil Engineering, 2021.

[32] J. Franke et al. Recommendations on the use of CFD in wind engineering, Proceedings of the International Conference on Urban Wind Engineering and Building Aerodynamics. 2004.

[33] Y. Ozmen, E. Baydar, and J. P. A. J. van Beeck. Wind flow over the low-rise building models with gabled roofs having different pitch angles, Building and Environment, Vol. 95, pp. 63–74, 2016.

View publication stats

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Electrochemical performance of NCA based cathodes for variable thickness of electrode through modelling and simulation

Pritesh Singh, Vaishali Dudeja, Amrish K. Panwar *

Department of Applied Physics, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Lithium-ion battery
NCA
Thickness
Polarization
Voltage
Capacity

ABSTRACT

In this study, modelling of lithium-ion battery (LIB) has been carried out using theoretical simulation. The simulation work has been performed for the better electrochemical performance of the battery with the $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ (NCA) type cathode material, Graphitic, Li_xC_6 as an anode material, and $LiPF_6$ as liquid electrolyte. Initially, the thickness variation of the cathode material of NCA has been performed. The cycle of the cells made of seven variable thicknesses of the NCA cathode has been estimated. Thereafter, the effect of optimized electrode thickness on the discharge curve of the cell at different C rates such as 0.5C, 1C, 1.5C has been obtained from the simulation results. The voltage drop at the cathode and anode ends due to the SEI layer has been simulated at the respective C rates for first (1^{st}) and 2000^{th} cycles. The results obtained from all three simulations show consistency with the previously reported literature based on experimental work. Hence, the cathode material (NCA) thickness is optimized and chosen easily according to the application and need.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

1. Introduction

Nowadays, the energies are generated by various non-renewable sources such as crude oil, coal, natural gases, etc. These sources pose a risk to the global atmosphere and climate change due to the production of CO_2 gases, air pollution, and greenhouse gases [1]. To overcome these issues, the research work is going on in this area to produce renewable energy sources such as solar cells, photovoltaic cells, micro-wind turbines, water energy, etc. [2]. But due to the unavailability of regular source of energy from these. There is utmost requirement of good energy storage devices to provide the constant supply to the electric consumers, so that supply cannot be affected by the fluctuation of the renewable energy sources. Now a days rechargeable batteries are one of the most useful energy storage alternates to fulfil the constant supply demand.

Rechargeable battery is an electrochemical cell with a positive electrode as cathode and a negative electrode as anode, are separated by a solid separator and filled with liquid electrolyte and follow the multiple charge – discharge. Among various rechargeable batteries, currently the lithium-ion battery (LIB) is most popular and useful technology in the battery sector. These batteries have high energy and power density, and better cycle capabilities than other types of rechargeable batteries. These qualities are very much important in the modern world and its applications like electrical and hybrid vehicles and one of the most important sectors i.e., energy storage systems which are used in renewable energy applications. The chemistry of the lithium-ion batteries follows the movement of lithium ions towards the positive end (cathode) through the organic electrolytic medium from the negative end (anode) during discharge process. Mostly, the anodes in lithium-ion batteries are made of graphite.

There are different types of cathode materials such as layered, $LiCoO_2$, spinel, $LiMn_2O_4$ (LMO), olivine, $LiFePO_4$ (LFP) and other compounds of similar structure [3,4,5]. Among the all types of developed cathode materials, the layered materials containing cobalt, nickel, and aluminum with different concentrations are of good capacity, good energy and power densities. As these different cathodes allow high lithium insertion and intercalation levels leading to large energy storage quantities. The lithium nickel cobalt aluminum oxides (NCAs) with variable composition of Ni, Co and Al content are a group of mixed metal oxides [6]. Some of them are important due to their application in lithium-ion batteries.

* Corresponding author.
E-mail address: amrish.phy@dtu.ac.in (A.K. Panwar).

https://doi.org/10.1016/j.matpr.2022.04.446
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

Please cite this article as: P. Singh, V. Dudeja and A.K. Panwar, Electrochemical performance of NCA based cathodes for variable thickness of electrode through modelling and simulation, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.446

NCAs are used as primary material on the positive electrode (cathode). The general formula is $LiNi_xCo_yAl_zO_2$ with × + y + z = 1. The rechargeable lithium-ion batteries consisting of NCA cathodes are currently available in the market, and also being used in the electric vehicles (EVs) and electric appliances with × ≈ 0.8. The batteries containing high Ni content in NCA cathode is referred as Ni-rich batteries. The voltage of those batteries lies between 3.8 V and 4.1 V, and has a nominal voltage of 3.6 V to 3.7 V. Hence, Ni-rich layered lithium transition-metal oxides, $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ (NCA) are archetypal cathodes used for EV batteries. Because of their high reversible capacity, good rate capability, and low cost. These materials have been studied extensively during the last two decades [7]. It has many applications such as in power grid, medical devices, and electric cars. Most notably Tesla, one of the leading manufacturers of electric vehicles, deploys NCA batteries in their electric vehicle's system. NCA batteries can store the same amount of energy as that of NMC batteries. Although NCA batteries are not widely used in the consumer industry, but are quite promising and used in the electric car industry.

The usable charge storage capacity of NCA is about 180 to 200 mAh/g, which is less than the theoretical storage capacity, 279 mAh/g [8]. However, the capacity of NCA is much higher than that of other alternative materials belonging to similar layered structure such as lithium cobalt oxide $LiCoO_2$ (LCO) with 148 mAh/g, lithium iron phosphate $LiFePO_4$ (LFP) with 165 mAh/g, and $LiNi_{0.33}Mn_{0.33}Co_{0.33}O_2$ (NMC 333) with 170 mAh/g. Due to its high voltage, NCA enables batteries with high energy density. Another advantage of NCA is its excellent fast charging capability, High specific energy, good specific power, long life span. However, NCA materials also have demerits in terms of safety issues as there is need of extra safety features and circuitry.

Hence, the theoretical studies have been attempted on NCA based cathode material in lithium ion-battery. Therefore, a model has been prepared through the simulation with defined initial parameters. A one-dimensional (1-D) model is prepared with the variation of the thickness of the NCA cathode. Finally, the optimized thickness has been estimated to show the best electrochemical performance and to investigate polarization effect, charge–discharge voltage etc.

2. Model geometry (1-D) of battery

Fig. 1(a) indicates the 1-D model geometry used in the modelling of lithium-ion battery without electrolyte. Here, the length starting from 0 to 5.5 represents the negative electrode i.e., Anode made of Graphitic Electrode, Li_xC_6 material. After 5.5 to 8.5, the geometry shows the length of the separator and the last part from 8.5 to 15.5 is the length of the Positive electrode (Cathode) material used as $LiNi_{0.8}Co_{0.15}Al_{0.05}O_2$ (NCA). Fig. 1(b) shows the whole cell geometry after dipped into the liquid Electrolyte, $LiPF_6$ and it is shown by blue colour.

In this model, the thickness of the anode and separator is kept fixed as 55 μm and 30 μm, respectively [9]. Whereas the thickness of the cathode is varied. The initial parameters and the defined values used for the simulation as anode, separator, cathode and other parameters are given in Table 1.

3. Results and discussions

Fig. 2 shows the simulated data of charge and discharge of the battery at different thickness of NCA cathode. The Fig. 2(a), represents the charge–discharge curve of the battery for a thickness of the cathode, 50 μm. The curve clearly shows that there is a hold between charging and discharging, and when the thickness is increased by 10 μm, the hold reduces to some extent which can be seen in the Fig. 2(b) where the thickness of the cathode is 60 μm. On further increment of the thickness by 10 μm i.e., making the thickness of the cathode 70 μm, the hold between charging and discharging has further decreased, but a few holding phenomena is still present when looked carefully depicted in Fig. 2(c). After taking the thickness to 80 μm, 90 μm, and 100 μm the hold is completely reduced but there is a sharp fall in the discharging voltage, as we can see in Fig. 2(e), (f), (g), respectively. Fig. 2(d) represents the charge and discharge curve of the battery in which the thickness of cathode is 75 μm, and it is clearly visible that the result is similar to that of the 80 μm thickness curve but there was a slight hold between the charge and discharge curve. These curves also depict that with increasing thickness the voltage IR drop is decreasing and has a constant voltage for longer duration.

Not only there is variation in the hold, but with the increment of thickness, the time taken in the charging process is also being reduced. It is clearly visible that, for the thickness of 50 μm the charging time is more than 3000 s and can be observed in Fig. 2 (a). Whereas, charging time for the battery having cathode of 80 μm thickness is less than 2000 s, Fig. 2(e), and if we see about the cathode of 100 μm thickness, the charging time is around 1500 s.

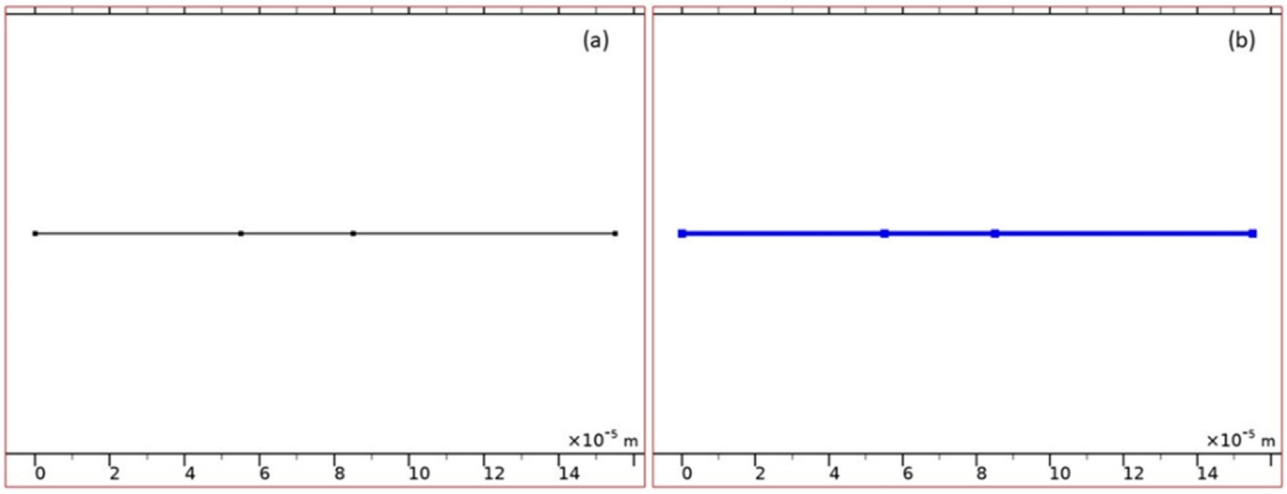


Fig. 1. (a) 1-D model geometry of anode, separator and cathode in battery, (b) 1-D model geometry of battery with electrolyte.

Table 1
Initial Parameters of the model.

Description	Value	Anode	Cathode	Separator
Particle radius negative electrode	–	$2.5 \times 10^{-6}\,m$	–	–
Particle radius positive electrode	–	–	2.5×10^{-7} m	–
Solid-phase volume fraction positive electrode	–	–	0.42	–
Electrolyte phase volume fraction positive electrode	–	–	0.41	–
Bruggeman coefficient for tortuosity in the positive electrode	–	–	2.98	–
Solid-phase volume fraction negative electrode	–	0.384	–	–
Electrolyte phase volume fraction negative electrode	–	0.444	–	–
Electrolyte phase volume fraction separator	–	–	–	0.37
Bruggeman coefficient for tortuosity in separator	–	–	–	3.15
Length of the negative electrode	–	5.5×10^{-5} m	–	–
Length of separator	–	–	–	3×10^{-5} m
Length of the positive electrode	–	–	7×10^{-5} m	–
Cell temperature	318.15 K	–	–	–
SEI layer conductivity	5×10^{-6} S/m	–	–	–
Initial SEI layer thickness	1×10^{-9} m	–	–	–
Number of cycles	2000	–	–	–

All these observations for variable thicknesses indicate that the thickness of about 80 μm of the cathode material could be optimize for better discharge capacity and minimize holding step during discharge for this thickness. The theoretical results have been compared to experimental data, and it reveals that for better results, the thickness of the cathode was chosen between 70 μm to 80 μm [10]. If it is considered specially for the load cycle at 75 μm thickness, apart from the hold between charging and the discharging, the charging time is more than 2000 s. Whereas for the thickness of 80 μm the charging time is less than 2000 s.

There is a voltage hold for sometimes while discharging at low thickness whereas with increasing thickness the voltage hold is reduced but there is a drop in the voltage while increasing the thickness of NCA cathode. The main reason we have found for this is the fact that there is a passivation layer formed which leads to high impedance, and as at low thickness, the number of ions produced are also be less as compared to the higher thickness materials [11,12]. The most of ions that are polarized may try to overcome the passivation layer, and its impedance leads to the hold of voltage, the rest of the polarized ions lead to voltage drop. Since there are a smaller number of ions, and most of them have been moved to overcome the passivation layer, the polarization is less that is why there is a low voltage drop. But with the increasing thickness, the number of ions produced increases which ultimately leads to heavy polarization, and this leads to more voltage drop. As there are more ions in thicker material, they overcome the passivation layer easily. The voltage drop due to polarization at different thickness of cathode is given in Table 2.

Thereafter, the discharge curve at different cycle rates and voltage drops due to SEI layer formation at different cycle rates for the thickness of 75 μm and 80 μm of cathode has been observed. Voltage in first cycle and 2000th cycle has been calculated at these two thicknesses with different cycle rate.

Fig. 3 shows the discharge curve and voltage drop due to SEI layer formation at different cycle rates, keeping the thickness of cathode material as 75 μm. Fig. 3(a) shows the discharge curve of the battery at 1st cycle and 2000th cycle, battery is being operated at 0.5C rate. The peak voltage obtained was around 3.8 V in the 1st and 2000th cycle and also there has been no change in voltage with time, as almost similar discharge profile is observed. Fig. 3 (b) shows the potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing of cycle number, there has been a significant drop in the potential at both the ends and also the amount of drop is similar at both the ends. The discharge curve of the battery at 1st cycle and 2000th cycle when operated at a 1C rate is shown in Fig. 3(c). The peak voltage obtained was around and slightly less than 3.8 V in the first and 2000th cycle but with time the voltage in the 2000th cycle is significantly less than the voltage in the first cycle, hence the difference is about 0.1 V. Fig. 3(d) indicates the potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing cycle number, there has been a significant drop in the potential at both ends, but after around 750 cycle numbers the potential drop at the negative electrode current collector end the potential drop is more than that at the negative electrode separator end, operated at 1C rate. The Fig. 3 (e) shows the discharge curve of the battery at 1st cycle and 2000th cycle when operated at a 1.5C rate. The peak voltage obtained was around 3.7 V in the first and 3.6 V in the 2000th cycle but with time, the voltage in the 2000th cycle is decreased by 0.2 V as compared to the first cycle. The Fig. 3(f) shows potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing cycle number, there has been a significant drop in the potential at both ends. The potential drop at both points is exactly equal with increasing cycle numbers. However, a slight change is visible after 1750 cycles.

Thereafter, the discharge curve and voltage drop due to SEI layer formation at different cycle rate at 80 μm thickness of cathode is shown. Fig. 4(a) shows the discharge curve of the battery at 1st cycle and 2000th cycle, battery is being operated at 0.5C rate. The peak voltage obtained was around 3.7 V in the first and 2000th cycle and also there has been no change in voltage with time, i.e., in the first and 2000th cycle, the discharge curve is almost the same. The Fig. 4(b) shows the potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing cycle number, there has been a significant drop in the potential at both the ends and also the amount of drop is similar and somehow equal at both the ends. The Fig. 4(c) shows the discharge curve of the battery at 1st cycle and 2000th cycle when operated at a 1C rate. The peak voltage obtained was around and slightly less than 3.7 V in the first and 2000th cycle but with time the voltage in the 2000th cycle is significantly less than the voltage in the first cycle, the difference is about 0.1 V. The Fig. 4(d) shows the potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing cycle number, there has been a significant drop in the potential at both ends, but after around 750 cycle numbers the potential drop at the negative electrode current collector end, the potential drop is more than that at the negative electrode separator end, operated at 1C rate. The Fig. 4(e) shows the discharge curve of the battery at 1st cycle and 2000th cycle when operated at a 1.5C rate. The peak voltage

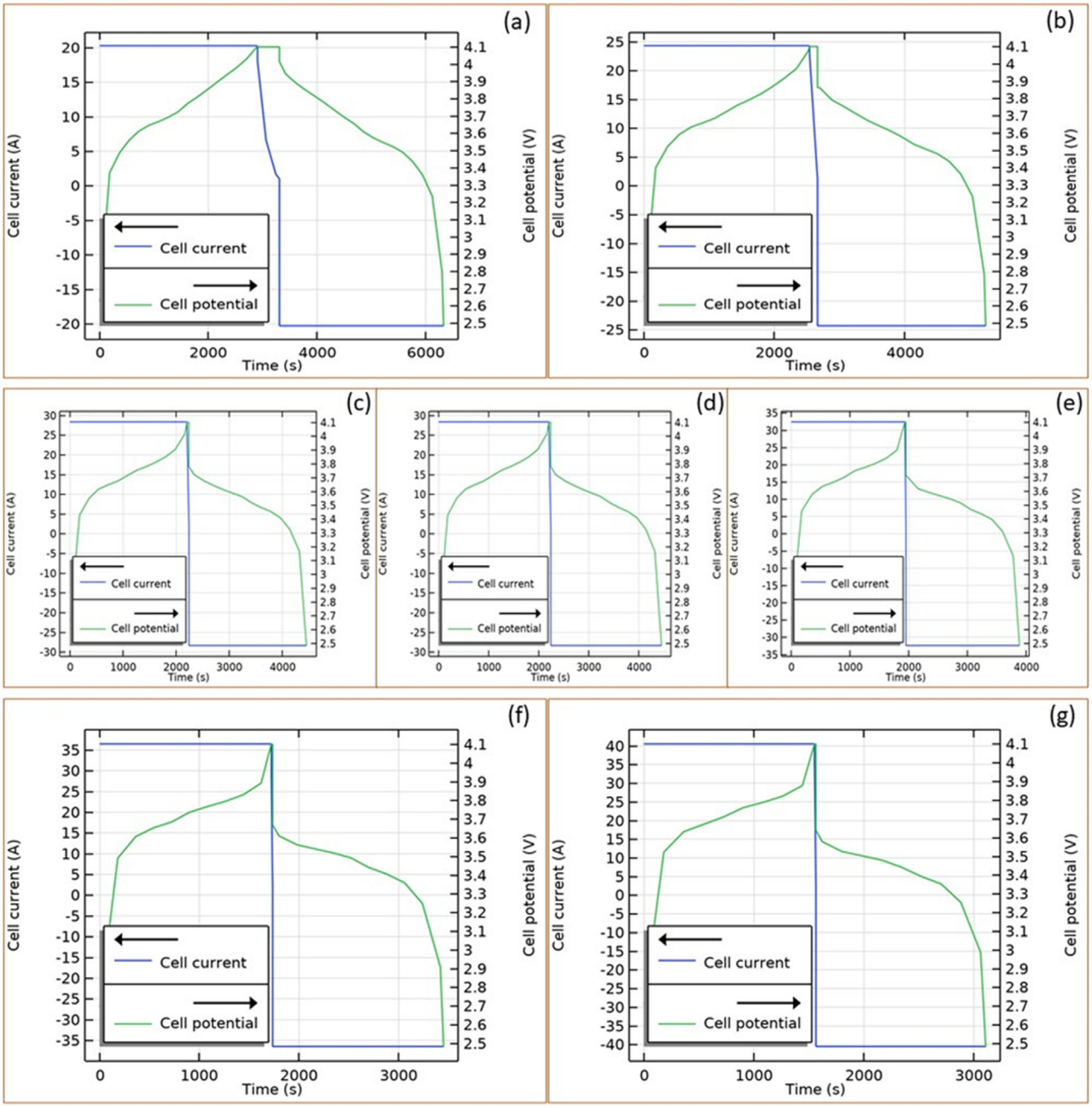


Fig. 2. Charge-Discharge curve of the battery at thicknesses of: (a) 50 μm, (b) 60 μm, (d) 75 μm, (e) 80 μm, (f) 90 μm, and (g) 100 μm thickness of NCA cathode.

Table 2
Voltage drop due to polarization at different thickness.

Thickness of cathode (μm)	Peak Voltage (V)	Nominal Voltage (V)	Voltage Drop ΔV
50	4.1	4.0	0.1
60	4.1	3.87	0.23
70	4.1	3.78	0.32
75	4.1	3.78	0.32
80	4.1	3.71	0.39
90	4.1	3.68	0.42
100	4.1	3.64	0.46

obtained was around 3.7 V in the first and 3.5 V in the 2000th cycle but with time, the voltage in the 2000th cycle is decreased by 0.2 V as compared to the first cycle. The Fig. 4(f) shows potential drop at the negative electrode current collector and negative electrode separator end, due to SEI layer formation. With the increasing cycle number, there has been a significant drop in the potential at both ends. The potential drop at both points is exactly equal with increasing cycle numbers. A slight change is visible after 1750 cycles. Voltage in first and 2000th cycles at different C rates has been calculated and given in Table 3.

4. Conclusion

The theoretical simulation of NCA based cathode has been carried out for variable thickness of the material under 1-D cell geometry. The charge–discharge curve at variable thicknesses starting from 50 μm to 100 μm of the NCA based cathode material are obtained. It is estimated that the thicknesses 75 μm and 80 μm

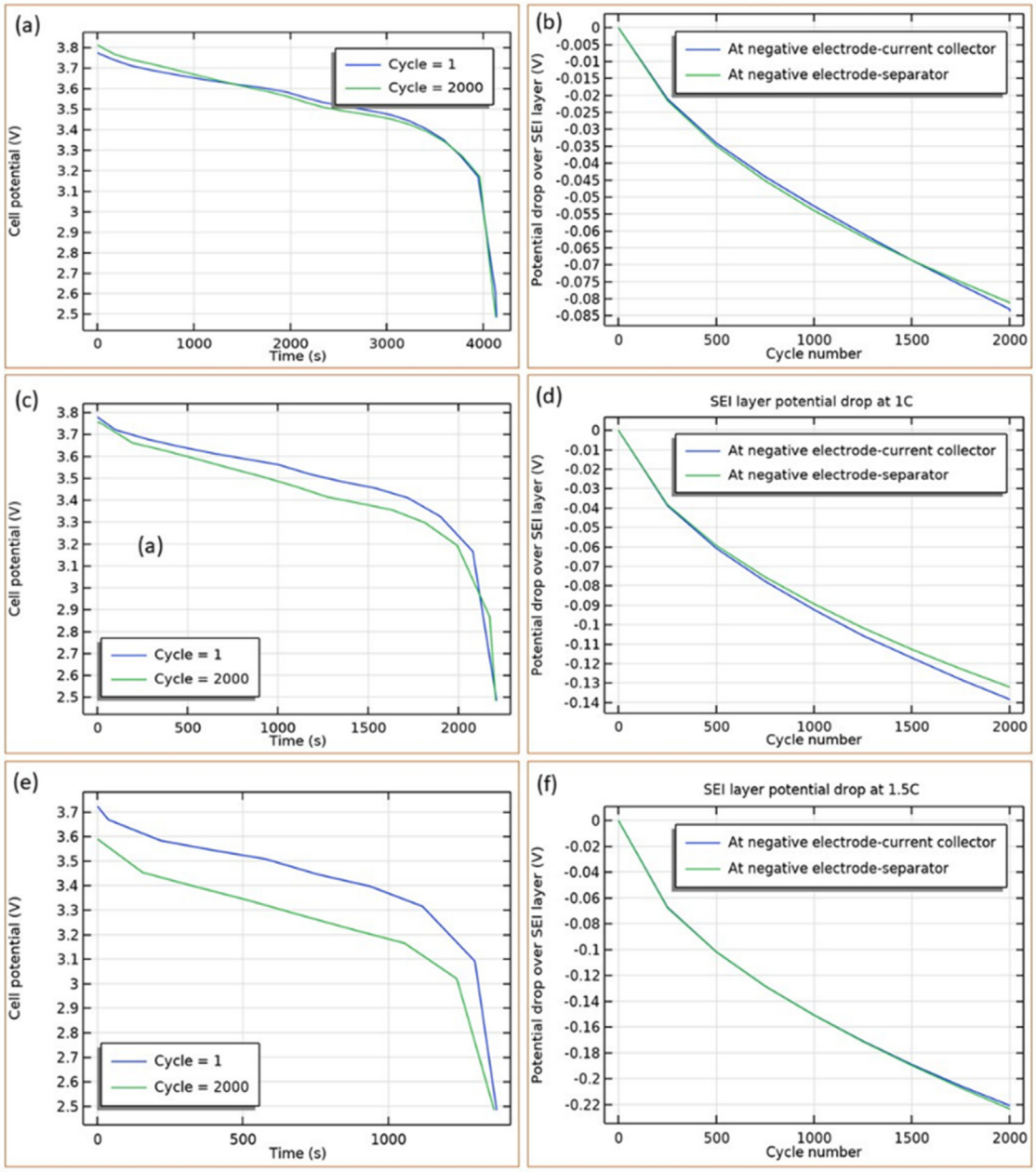


Fig. 3. Discharge curve and voltage drop due to SEI layer at (a, b) 0.5C rate; (c, d) 1C rate, and (e, f) 1.5C rate, for NCA cathode thickness of 75 μm.

of NCA are found to be best for higher discharge time/ discharge capacity and relatively lower polarization effect comparing with higher thicknesses than 80 μm. Further, simulated results were obtained for three different cycle rates: 0.5C, 1C, and 1.5C for these two thicknesses of 75 μm and 80 μm. Therefore, the discharge curve was obtained for 1st and 2000th cycles for both the thicknesses. Hence, it is observed that the polarization is increased with cycling up to 2000th cycle and it increases with increasing the discharge rates for both the thicknesses of 75 μm and 80 μm. Moreover, the effect on potential drop is by and large similar due to SEI layer and it is more prominent to 80 μm. Overall with these observations, it is concluded that NCA cathode with a thickness between 70 μm to 80 μm performs highly well and may give better electrochemical performance.

CRediT authorship contribution statement

Pritesh Singh: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **Vaishali Dudeja:** Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **Amrish K. Panwar:** Supervision, Formal analysis, Visualization, Writing – review & editing.

Declaration of Competing Interest

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

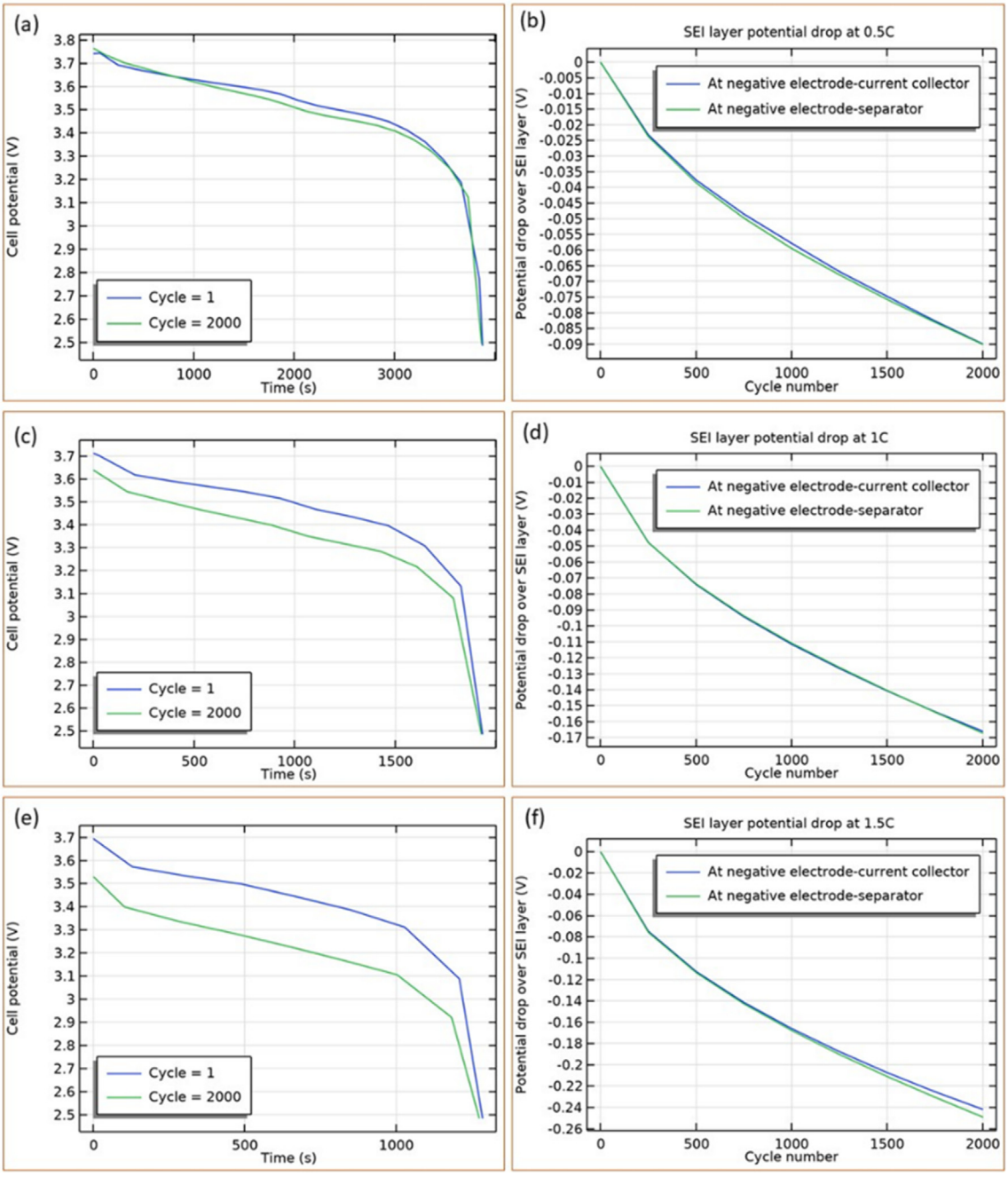


Fig. 4. Discharge curve and voltage drop due to SEI layer at (a, b) 0.5C rate, (c, d) 1C rate, and (e, f) 1.5C rate, for NCA cathode thickness of 80 µm.

Table 3
Capacity at different cycle rate and nominal voltage at first and 2000th cycle.

Thickness	C rate	Discharge Capacity (mAh/g)	Voltage in 1st Cycle	Voltage in 2000th Cycle
75 µm	**0.5**	180.30	3.81	3.78
	1	193.40	3.78	3.76
	1.5	180.03	3.72	3.59
80 µm	**0.5**	191.34	3.75	3.75
	1	171.17	3.70	3.64
	1.5	176.61	3.70	3.52

Acknowledgements

The authors are thankful to Delhi Technological University for providing financial support through the project grant: F. No. DTU/IRD/619/2019/2114.

References

[1] A.M. Omer, Renew. Sustain. Energy Rev. 12 (2008) 2265–2300.
[2] F. Calise, Francesco Liberato Cappiello, Massimo Dentice d'Accadia, and Maria Vicidomini, Renew. Energy 160 (2020) 633–652.
[3] G. Singh, R. Thomas, R.S. Arun Kumar, A.M. Katiyar, J. Electrochem. Soc. 159 (4) (2012) A470.
[4] R. Saroha, A. Gupta, A.K. Panwar, J. Alloy. Compd. 696 (2017) 580–589.
[5] R. Saroha, A.K. Panwar, Y. Sharma, P.K. Tyagi, S. Ghosh, Appl. Surf. Sci. 394 (2017) 25–36.
[6] H.-H. Ryu, N.-Y. Park, J.H. Seo, Y. Young-Sang, M. Sharma, R. Mücke, P. Kaghazchi, C.S. Yoon, Y.-K. Sun, Mater. Today 36 (2020) 73–82.
[7] G.W. Nam, N.-Y. Park, K.-J. Park, J. Yang, J. Liu, C.S. Yoon, Y.-K. Sun, ACS Energy Lett. 4 (12) (2019) 2995–3001.
[8] N. Nitta, W. Feixiang, J.T. Lee, G. Yushin, Mater. Today 18 (5) (2015) 252–264.
[9] B. Wang, L. Le Fevre, A. Brookfield, E. Mcinnes, R. Dryfe, Angewandte Chemie Int. Ed. (2021).
[10] M. Singh, J. Kaiser, H. Hahn, J. Electrochem. Soc. 162 (7) (2015) A1196.
[11] J.B. Goodenough, K.-S. Park, J. Am. Chem. Soc. 135 (4) (2013) 1167–1176.
[12] M. Broussely, F. Ph Biensan, B.P. Blanchard, S. Herreyre, K. Nechev, R.J. Staniewicz, J. Power Sources 146 (1–2) (2005) 90–96.

JMEPEG
https://doi.org/10.1007/s11665-022-06877-3

©ASM International
1059-9495/$19.00

TECHNICAL ARTICLE

Enhancement of Microstructural and Mechanical Properties of Ultrasonic Vibration-Assisted Cold Metal Transfer Welding of 304 Stainless Steel

N. Yuvaraj, Jayanta Ghosh Roy, and Vipin

Submitted: 20 December 2021 / Accepted: 4 March 2022

Ultrasonic vibrations aid fabrication techniques that enhance the structural efficiency of manufactured products. It gives an appreciable advantage in the microstructural and mechanical properties of products. In this paper, Ultrasonic assisted Cold Metal Transfer Welding (U-CMT) technique used to join the AISI 304 stainless steel. The welded joints are analyzed using optical microscopy, SEM, and XRD. The microhardness, tensile, and residual testing of the U-CMT welded joint is evaluated and compared with CMT welded joint. The microhardness and tensile results of the U-CMT welded joint show considerable improvement in their mechanical properties due to the refinement of grains in the fusion region, which is produced by ultrasonic vibrations. With vibration welded samples, results show a reduction of residual stresses compared to those without vibration welded samples. This study chose two different ultrasonic vibrational amplitude 50 and 99 μm and three different welding currents, 85, 90, and 95A. U-CMT welded joint prepared with a welding current of 95A and ultrasonic vibrational amplitude of 99 μm shows 100% weld joint efficiency.

Keywords AISI 304, CMT, residual stresses, tensile strength, ultrasonic vibration

1. Introduction

Austenite stainless steel has been extensively used in structural, automotive, medical, railways, dairy, and food-processing owing to its availability, economic, and ease of manufacturing. These stainless steels have an attractive mix advantage of physical and mechanical properties (Ref 1, 2). Conventional welding processes such as MIG/MAG, SAW, and SMAW weaken the stainless steel welded joint properties due to high heat input, which results in solidification induced defects. Controlling constant heat input is difficult by these welding processes (Ref 3, 4). Heat input is an important parameter to achieve sound weld strength. The cold metal transfer welding (CMT) process was developed based on MIG welding, which has a short circuit mode for controlling the wire's retraction movement owing to the detachment of liquid metal droplets from the filler wire. The current is maintained at a minimum during the short circuit mode due to the low heat developed during welding. CMT welding technique is competent to produce negligible distortion with spatter-free thin sheet joints of different materials with lesser heat input, low dilution rate, and minimum residual stresses (Ref 5, 6). In recent years, ultrasonic-assisted welding has given more attention due to grain refinement of welded material and increased homogeneity of filler material with the base material (Ref 7, 8). The ultrasonic vibration treatment during welding is an alternative method for heat treatment and eliminating the post-weld heat treatment of the welded joint to lower the residual stresses for enhancement of mechanical properties (Ref 9). The different methods of ultrasonic probe located by the various researchers during welding are from the side of the substrate, top of the substrate, to the filler metal, to the electrode, and directly on the weld pool for achieving the best outcome (Ref 10-12). The ultrasonic vibration was introduced directly on the welding pool through filler metal for TIG welding of ferrite stainless steel by Watanabe et al. (Ref 13) and found that tensile strength and elongation of the welded joint are improved due to more equiaxed grain formed in welded joint in comparison to samples processed without vibration welded joint samples. Xie et al. (Ref 14) imposed ultrasonic vibrations on the MIG arc to fabricate Al-Zn-Mg alloy welded joint and revealed that larger HAZ and fusion region sizes were observed in vibration-assisted welded joints than conventional MIG welded joints. Dai (Ref 15) positioned the ultrasonic probe on the top surface of Al 7075-T6 alloy for TIG welding, enhancement of hardness in fusion region, and better weld penetration observed ultrasonic-assisted welding in comparison to samples processed without vibrations welded samples.

Zhao et al. (Ref 16) inserted a tungsten needle with ultrasonic vibrations in the front side of the molten pool during welding. Grain refining occurs at welded material due to the tungsten needle being located at the low-temperature zone of the molten pool material. Chen et al. (Ref 17) reported that ultrasonic vibrations promoted heterogeneous nucleation of Al-Li alloy during TIG welding. Yuan et al. (Ref 18) placed an ultrasonic probe at the back of the arc to stir the weld pool

N. Yuvaraj, Jayanta Ghosh Roy, and **Vipin**, Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India. Contact e-mails: yuvraj@dce.ac.in, jghoshroy@gmail.com, vipin@dce.ac.in.

Published online: 09 May 2022

during welding. The vibrations are transferred to the molten metal, which results in refined grains of welded material. Tian et al. (Ref 19) placed an ultrasonic peening system on the Al alloy workpiece with CMT welded and found that refinement of grains with lesser porosity of welded joints. Fattahi et al. (Ref 20) investigated TIG welding of Al alloy with the inclusion of nanoparticles and ultrasonic vibrations under the workpiece and found that with enhanced mechanical properties of welded joints. Ning and Cong (Ref 21) employed SS powder on low carbon steel to fabricate stainless steel parts by applying vibrations in the laser engineering net shaping technique (UV-LENST). The mechanical properties of the parts produced by UV-LENST are much higher than without ultrasonic vibration fabricated samples due to finer dispersion of particles, refinement of grain size, lower residual stress, and lower micro-cracks with reduction of porosity. Chen et al. (Ref 22) introduced the ultrasonic vibration in pulsed TIG of Q235 steel and found that refinement of grains occurred in weld seams compared to conventional TIG. With the application of ultrasonic vibration assisted various welding techniques such as MIG (Al 1060 and Al 1003 filler wire (Ref 23), MIG (Al-Cu) (Ref 24), MIG (Al 1060 and S301 filler wire) (Ref 25), TIG (304 SS and FM-52M filler wire) (Ref 26), TIG (316L and L415 dissimilar) (Ref 27), Plasma (304 SS) (Ref 28), LASER (A 516-55 low carbon steel) (Ref 29), (AISI 321 SS) (Ref 30), (Hastelloy C-276 and 304 SS) (Ref 7), (Low carbon steel) (Ref 31), (Inconel 718 superalloy) (Ref 32), (Al5052) (Ref 33) and FCAW (H40 Marine steel) (Ref 34) are investigated by various researchers.

Residual stresses are developing in many fabricated welded joints. These are measured by destructive and non-destructive techniques. The various destructive methods such as a deep hole, ring-core, hole-drilling, and contour techniques are commonly used. In non-destructive types, x-ray diffraction, Barkhausen noise, Ultrasonic, and Neutron diffraction types are widely used (Ref 35, 36). Residual stresses of welded joints are evaluated by the blind hole-drilling process (Ref 37, 38) and deep hole contour method (Ref 39) by various researchers. With the application of μ-X360 and iXRD x-ray diffractometer, the residual stresses of welded carbon steels are investigated by Lai and Wu (Ref 40). The entire region of the stress and inter-planer atomic spacing on the selected area can be measured accurately by the x-ray diffraction method. Kumar et al. (Ref 41) studied the residual stresses of laser-welded austenite stainless steel using the x-ray diffraction Prototype iXRD measurement method. With the reduction of heat input from 75 to 45 J/mm during the welding, the residual stresses of welded joint change from tensile to compressive nature. Gadallah et al. (Ref 42) measured the residual stresses of medium carbon steel welded joint by linear friction welding by x-ray diffraction and contour technique and found that the applied welding pressure changes the variation of residual stresses behavior by both the approaches. Lesyk et al. (Ref 43) modified the stainless steel surfaces by various surface modification techniques and studied the residual stress of the subsurface layer by the x-ray diffraction method. Water jet cavitation peening modified surface has the lowest residual stresses, suitable for wear-resistant and anti-corrosive applications. Vykunta Rao et al. (Ref 44) investigated the effect of vibration with different vibrometer input voltage (50–230 V at an interval of 10 V) during TIG welding of Al5052 alloy. With increased input voltage up to 160 V, the strength of the welded joint is enhanced due to reduction in grain size; with a further increase in input voltage, mechanical properties deteriorate as a result of grain size increase. Limited work was dedicated to studying the application of ultrasonic-assisted CMT welding of austenite stainless steel. This paper selects three different welding currents to fabricate welded joints; with ultrasonic vibrational assistance and without vibrational assistance. Two different ultrasonic vibrational amplitude are chosen to investigate microstructural effects, microhardness, tensile properties, and residual stresses of welded joints.

2. Materials and Methods

AISI 304 SS rolled sheet with a size of 100 mm × 60 mm × 2 mm was butt welded with 1.2 mm diameter ER 308L SS filler. The composition of base material and filler wire is listed in Table 1. TPS 400i CMT welding machine was employed to conduct the experiments. Selected welding parameters are exhibited in Table 2. Constant welding speed of 4 mm/sec and contact to workpiece distance of 7 mm has been fixed after a vast number of trial experiments. With the selected input parameters, the required voltage in the range of 10-20.8 V is chosen automatically by the machine. Pure argon (99.99%) shielding gas with a 14 L/min constant flow rate was used to conduct all experiments. The ultrasonic probe is clamped on the fixture to touch the middle of the two workpieces from the width side, shown in Fig 1(a). Ultrasonic vibrations of 20 kHz constant frequency and vibration amplitude in the range of 11-99 μm are generated by the ultrasonic generator. For microstructural examination, the cross-sectional specimens are extracted from the transversal direction of the welded joint. The specimens are polished as per the standard metallographic approach and etched by 30 ml HCl + 120 ml distilled water + 10 g $FeCl_3$ (Ref 45) solution. The optical microscope (Make OLYMPUS) was employed for macro- and microstructural studies. A microhardness tester measured the microhardness of the cross section of the welded joint samples at a load of 200 gm with a dwell period of 20 seconds. Cross-tensile specimens are extracted in the transversal direction from welded joint as per ASTM: E8/E8M-011 standards by wire EDM. Tensile tests are conducted using Tinius H50KS universal testing machine at a 1 mm/min ram speed. Figure 1(b) shows the extraction of Tensile and hardness specimen from welded joints, (c) shows the tensile test specimens, and (d) shows the tensile test specimen dimensions. Residual stresses are computed by the Cosα method using x-ray diffraction method full 2D-based (PULSTEC μ-X360n) high-resolution machine. The equipment is equipped with a standard chromium x-ray tube with a 1 mm diameter of the collimator and 30 kV and 1 mA specifications.

3. Discussions and Results

3.1 Microstructure

The fabricated welded joints are visually examined before the characterization and ensuring that the welded samples are free from surface defects. Figure 2 illustrates the schematic diagram of weld metal geometry. The variation in macroimages of weld metal geometry of CMT and U-CMT welded samples is shown in Fig. 3. The U-CMT sample's weld width is enlarged

Table 1 Chemical composition of base material and filler wire (% weight)

Materials	Fe	Cr	Mn	Ni	Si	C	P	S	Mo	N	V	Ti
AISI304 SS	Balance	17.90	1.88	8.38	0.25	0.043	0.046	0.016	0.224	0.05	0.07	0.006
ER308L SS	Balance	19.15	1.42	10.03	0.42	0.029	0.022	0.012	0.256	0.01	0.03	0.003

Table 2 Selected welding parameters and welded joint results

Exp. No.	Weld type	Current, A	Ultrasonic vibrational amplitude, μm	UTS, MPa	Elongation, %	Joint efficiency, %	Fusion region Microhardness, HV
Sample 1	U-CMT	85	50	679	95.8	93.8	175.8
Sample 2	U-CMT	85	99	698	107	96.5	178.2
Sample 3	U-CMT	90	50	657	95.3	90.8	170.3
Sample 4	U-CMT	90	99	681	91.1	94.2	172.7
Sample 5	U-CMT	95	50	688	102	95.1	185.6
Sample 6	U-CMT	95	99	725	113	100	188.5
Sample 7	CMT	85	…	666	95	91	164.2
Sample 8	CMT	90	…	625	90.5	92	158.6
Sample 9	CMT	95	…	717	116	99.4	180.7

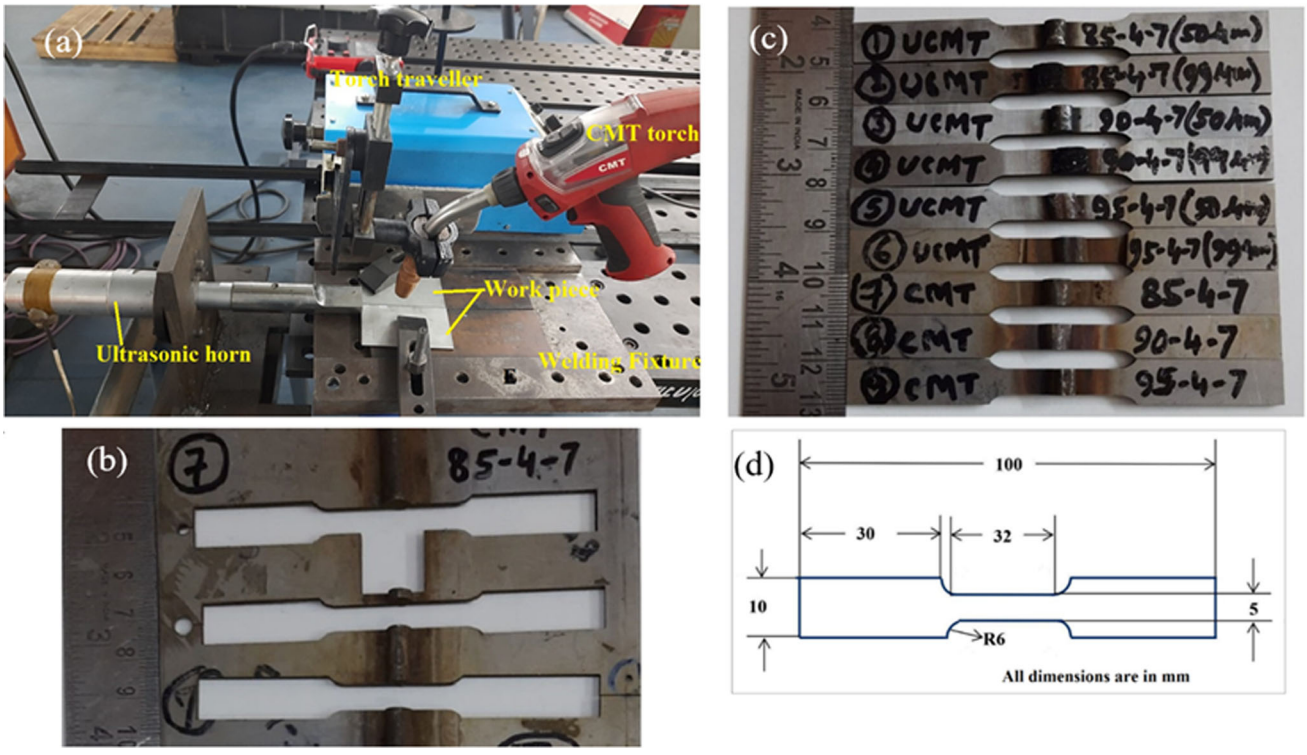


Fig. 1 (a) Experimental set-up (b) Tensile and hardness specimen extracted from the welded joint (c) Tensile test samples (d) Tensile test specimen dimensions

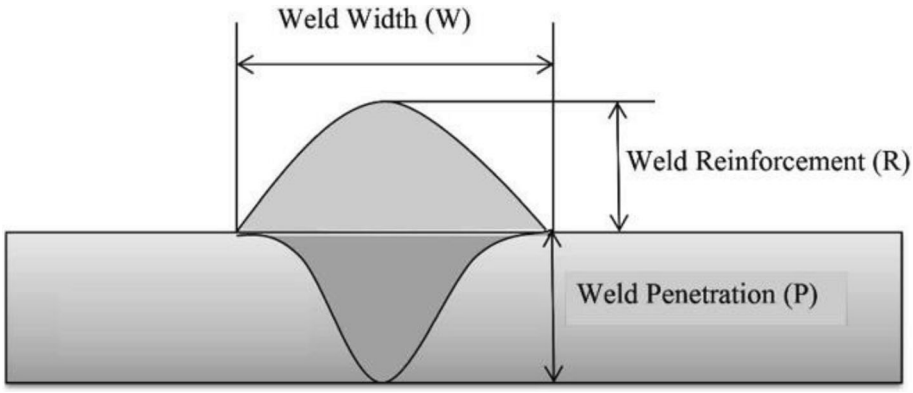


Fig. 2 Schematic diagram of weld metal dimensions

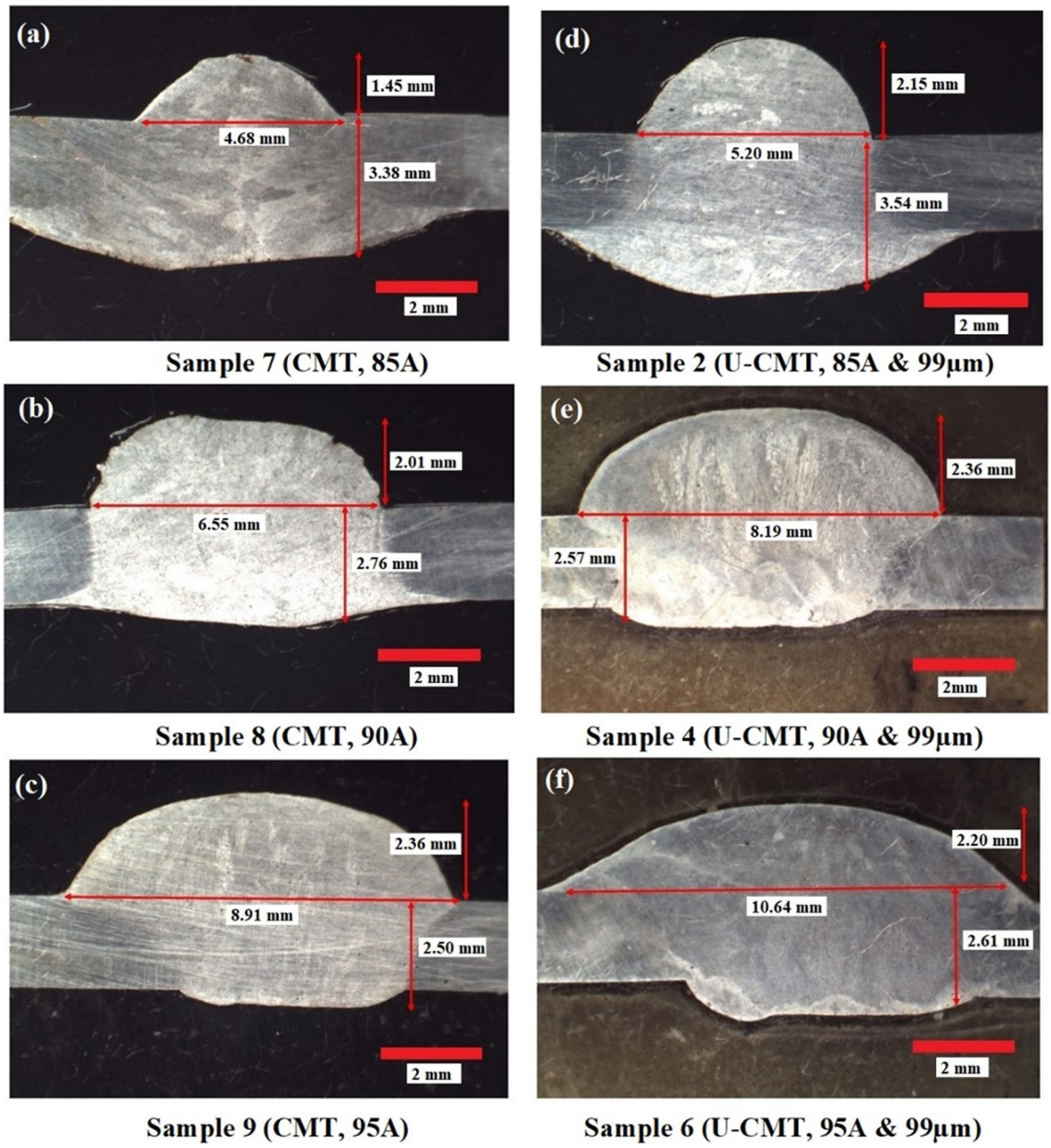


Fig. 3 Variation in macroimages of weld metal geometry of CMT and U-CMT welded samples

compared to the CMT sample, which is seen in macroimages of welded samples. The microstructure of the fusion region, HAZ and base metal of CMT, and U-CMT welded samples are shown in Fig. 4 and 5, respectively. Sun et al. (Ref 46) reported that the weld width of ultrasonic-assisted 304 SS TIG welded joints increases in comparison with non-vibration-assisted welded joints. The arc vibration energy increases the effect of push force continuously in the molten arc due to more filler metal being fused with the base metal. In the HAZ of the CMT welded sample (Fig. 4c), long, size dendrites and columnar shape grains were observed. With vibration, the smaller size dendrites and equiaxed grains were observed, which is seen in Fig. 5(c). With vibration, the molten weld material can break the dendrites due to the nucleation site being formed. During ultrasonic vibration-assisted welding, the temperature distribution is uniform (Ref 47). Wei (Ref 48) reported that the vibration fastens the disperse of molten weld material uniformly and adjusts the weld pool temperature uniformly as a result of equiaxed grains formed. The ultrasonic vibration introduction in weld pool material causes the cavitation effect in welded material, resulting in the finer size of grains of primary δ ferrite and an increase in dendritic fragmentation (Ref 18, 22). Further vibration boosts the solid–liquid interface with the help of supercooling to nucleate from weld pool liquid. Skeletal and lathy ferrite is developed in the austenite matrix in the weld region, which is seen in Fig. 4 and 5. In HAZ, dendrite and inter-dendritic phases and a few δ ferrite patches also are observed. The lower input vibrational amplitude has a better degree of grain fragmentation. With an increase in welding current, there will be increase in grain fragmentation; subsequently, the grain size of welded samples is decreased. Chen et al. (Ref 49) reported that the ultrasonic amplitude has an appreciable effect on grain fragmentation of ultrasonic-assisted welding of Al. The linear relationship takes place between welding current and grain fragmentation. The nonlinear relationship takes place between ultrasonic amplitude and grain fragmentation. Figure 6 shows the microstructure of HAZ and fusion regions of U-CMT samples welded at 50 μm and 99 μm vibrational amplitude. With an increase in vibrational amplitude, smaller size ferrite (δ) was observed in the fusion region. FESEM images (Fig. 7) at the weld metal show the austenite (γ) at the primary region and ferrite (δ) at grain boundaries. During solidification of welded material, delta ferrite and austenite grains formed with no trace of equiaxed grains. The δ ferrite and austenite grains form due to peritectic–eutectic reactions and moderate cooling rate during solidification. Pandey (Ref 50) also observed the δ ferrite formation in the 304 SS welded joint due to a lack in phase transformation.

The Ni-enriched dendrite phase is austenite, and Cr-enriched inter-dendritic phase is ferrite in the fusion region (Ref 51). EDS analyses have been carried out in austenite (γ) and ferrite (δ) phases of CMT and U-CMT samples to analyze the content of main alloying elements. The points taken for EDS studies of CMT and U-CMT samples are shown in Fig. 7(b) and (d), respectively. Figure 8 shows the EDS analysis of welded samples. Table 3 exhibits the EDS results % w of Cr, Ni, and Fe contents in the CMT and U-CMT samples. With the employment of vibration, the difference in Cr content in the austenite and ferrite region is reduced compared to without vibration sample. In addition, the Cr content in the ferrite region of the U-

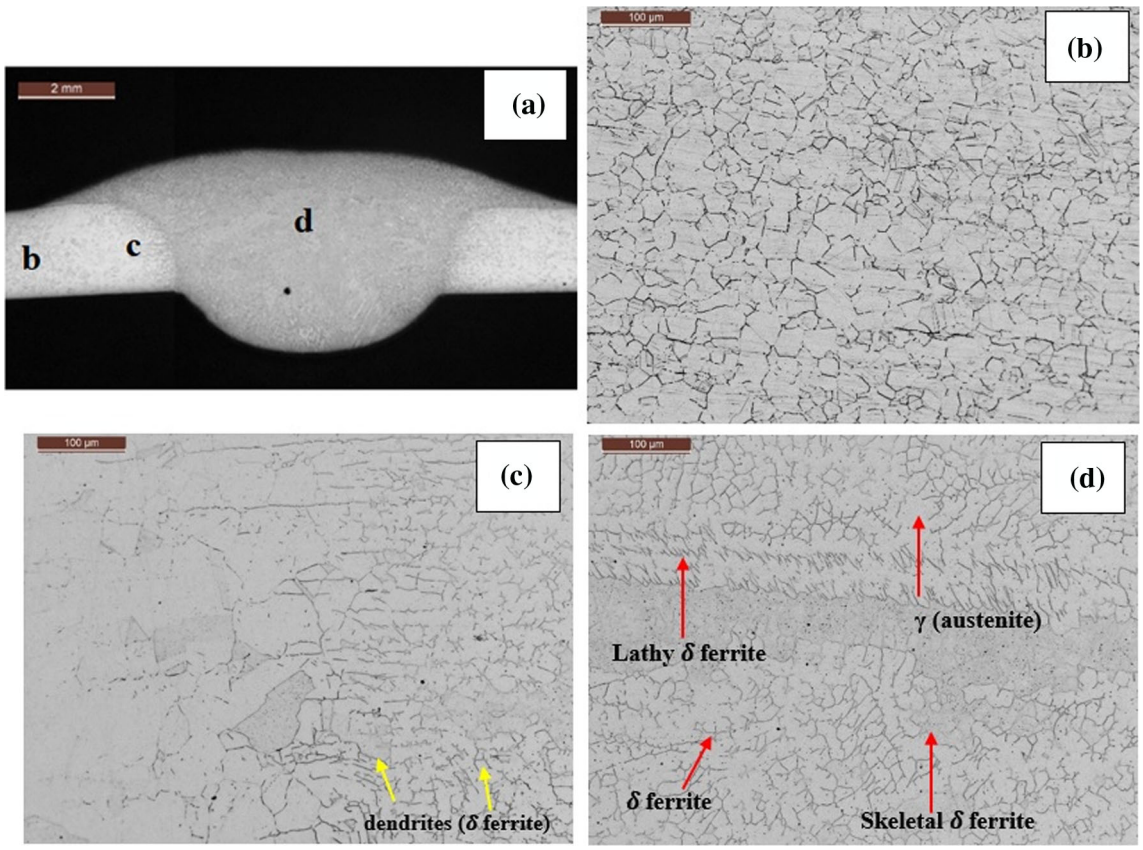


Fig. 4 Macrostructure and (b) microstructure of Base metal (c) HAZ (d) Fusion region of welded sample 9 (CMT, 95A and 99 μm)

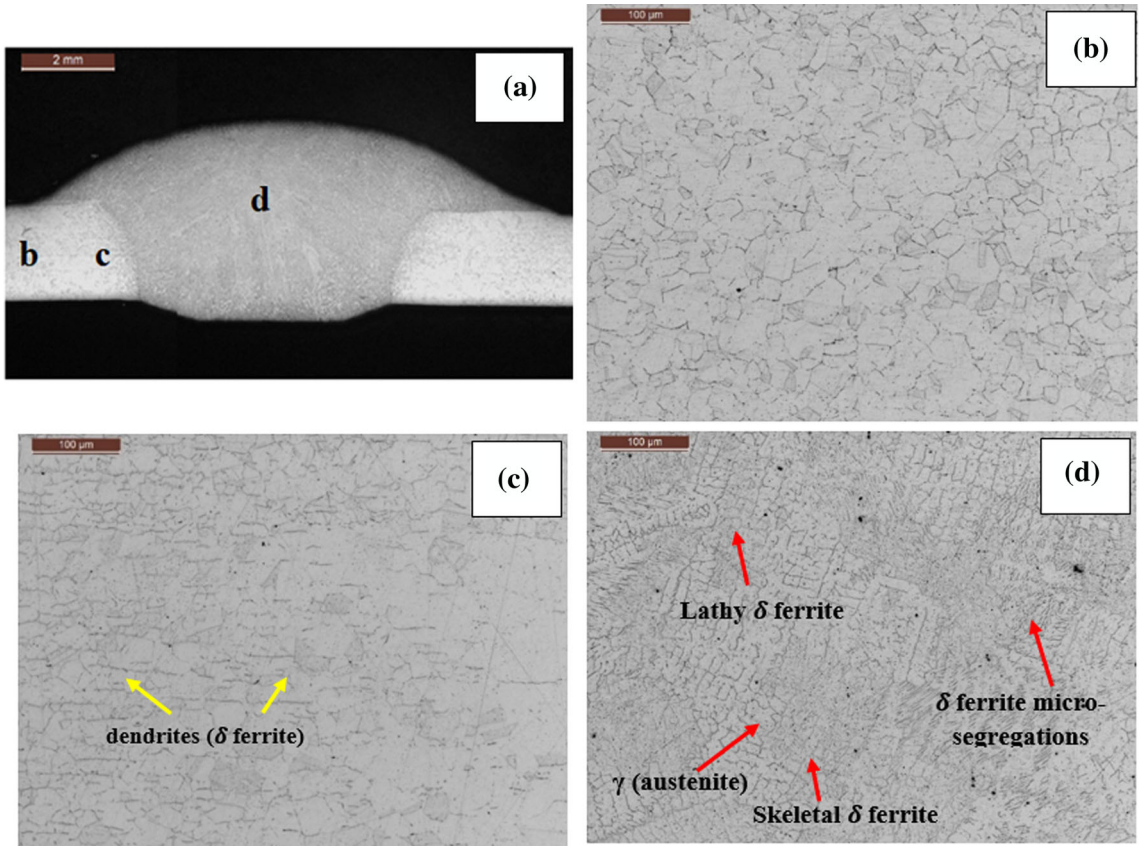


Fig. 5 (a) Macrostructure and (b) microstructure of Base metal (c) HAZ (d) Fusion region of welded sample 6 (U-CMT, 95A and 99 μm)

CMT sample is higher than the CMT sample. A similar trend of Ni content was also seen in the austenite region. The chances of Cr and Ni elements may be translated from filler wire to fusion region. The results revealed that ultrasonic vibration has a remarkable effect on the uniform distribution of alloying elements and refined microstructure. Figure 9(a) and (b) shows the grain size distribution chart for CMT and U-CMT samples of HAZ and fusion region, respectively. The average grain size is measured using Image J analysis from the microstructure. The selected base material's average grain size is around 50 μm. The average grain size of the U-CMT samples in the fusion region is approximately 36 μm, and for CMT, samples are about 45 μm. The vibrations reduces the size of grains and contact time between the grains during solidification of the molten weld pool, which results in the reduction of micro-segregation in the fusion region, which is shown in Fig. 5(d). In addition, enhanced Cr, Ni, and Fe contents segregated surface is formed in the dendritic and inter-dendritic regions. Figure 9 shows the XRD pattern of base material and welded samples. In base material (Fig. 10(a)), only austenite peaks were detected, and ferrite phases not detected due to smaller quantity may be present in the material, which XRD is not capable of detecting. On the other hand, weld material consists of Austenite (Fe, C) and Ferrite phases. During solidification of the molten weld pool, the ferrite solidifies in the fusion region, and then, transformation of ferrite to austenite takes place (Ref 27, 45). In the low heat input CMT process, the weld material cools faster and has insufficient time to complete the phase transformation as a cause of ferrite retained in the fusion region, which is seen in Fig. 10(b) and (c). In the welded samples, the intensity of the γ austenite plane has higher Cr enriched, and Ni depleted phases are formed (Ref 52). With the employment of ultrasonic vibrations during welding, the different concentrations of Cr and Ni occur between these elements, and inter-diffusion occurs in ferrite and austenite phases. In the CMT welded samples, $Fe_{0.64}$ $Ni_{0.35}$, $Cr_{0.19}Fe_{0.7}Ni_{0.11}$, and $C_{0.09}Fe_{1.191}$ phases are detected. In U-CMT, welded samples $Fe_{0.64}$ $Ni_{0.35}$, $Cr_{0.19}Fe_{0.7}Ni_{0.11}$, and $Cr_9Fe_{42}Mo_2Ti_5$ are detected.

3.2 Micro hardness

The microhardness evaluation of the CMT and U-CMT welded samples taken at the fusion region is exhibited in Table 2. Base material hardness is around 210HV. The increasing trend of microhardness value is observed in the order of fusion region, HAZ, and base metal in welded samples, as seen in Fig. 11. Minimum hardness was observed in the fusion region of all-welded samples. During welding, the heat causes the annealing effect in the fusion region, and recovery takes place for lesser hardness. In the HAZ, the tendency of hardness occurs due to a slow cooling rate in the fusion line; further grain growth occurs. However, near the base metal, a faster cooling rate occurs; therefore, finer grain may occur in the base metal (Ref 54). The hardness of U-CMT samples increases due to the refinement of grain size and decrease in δ-ferrite in the fusion region. In 304 SS weld material, Cr and Ni contents get atomic mobility due to increased ultrasonic vibration. Ni is quickly diffused in the austenite matrix and decreases in δ-ferrite phase. Hsieh et al. (Ref 55) reported a similar type of results for TIG welding of 304 SS with vibration. In the fusion region of U-CMT samples, grain

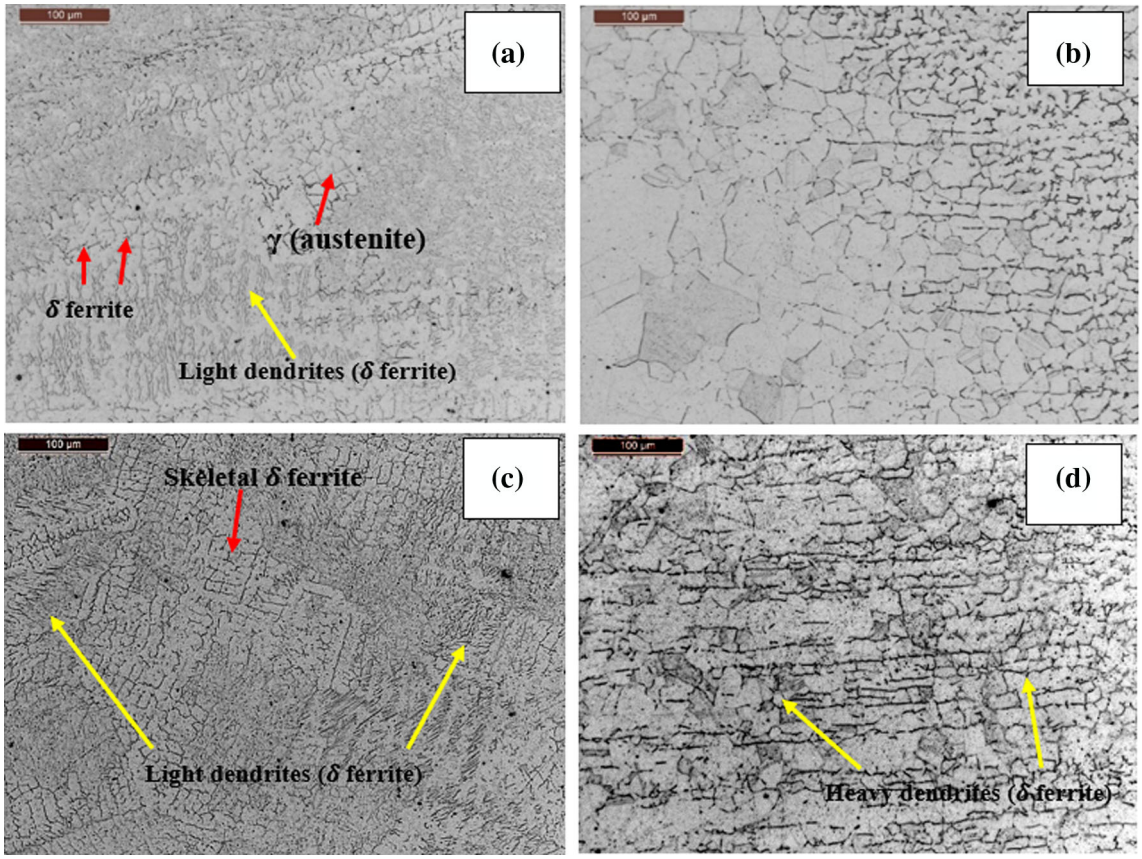


Fig. 6 Microstructure of (a) Fusion region and (b) HAZ of sample 3 (U-CMT, 90A and 50μm) and (c) Fusion region and (d) HAZ of sample 4 (U-CMT, 90A and 99 μm)

refinement occurs due to ultrasonic vibration breaking the columnar grains for enhancement of hardness (Ref 56). The ultra-sonication action generates many split dislocations in a hot molten weld pool which experiences higher hardness due to the work hardening effect of weld metal (Ref 30). Kolubaev et al. (Ref 29) reported that with the employment of vibrations, the grains of the steel welded joint are refined; as a result, a higher hardness is observed in the joint.

3.3 Tensile Properties

Typical tensile stress–strain graphs of welded samples are shown in Fig. 12. Base material tensile strength and elongation are 723.3 MPa and 60%, respectively. The more significant tensile strength of U-CMT samples welded with 95A and vibrational amplitude of 99 μm is 725 MPa, which is the marginal strength enhancement, and there is a significant improvement in elongation in comparison with CMT samples. In U-CMT samples, the strain gap is smaller inside the refined grains, and the grains form a uniform distribution, leading to lower fracturing of grains compared to coarser grains. The refined grain size welded joint undergoes extensive deformation before it is broken and attains good elongation (Ref 27). Welding current, speed, and contact to workpiece distance (CTWD) influence welding parameters for weld strength properties due to the evolution of appropriate heat input and deeper weld penetration (Ref 57). The tensile test results are presented in Table 2. The reduction of weld strength material due to higher current generates more material melting. The fusion region material experiences smaller dendrites and lesser inter-dendrite regions at appropriate welding current (Ref 58). At a higher current, the generation of heat input during welding is to a greater extent, the welded material undergoes an increase in grain coarsening effect, which deteriorates the tensile properties. In addition, lower tensile strength joint experiences brittle intermetallic compounds in the fusion region and more extensive dendrites formation in HAZ. The CMT welded samples experience lower tensile properties due to the formation of larger size dendrites and more inter-dendrite spacing in the fusion region. The broken tensile fractured surface of CMT and U-CMT tested samples FESEM micrographs is exhibited in Fig. 13. Various sizes and shapes with dimples were observed in the fractured surfaces, which revealed that the considerable fracture mechanism was ductile. A similar dimpled fractography with substantial precipitation was followed in welding 304 SS by Yan et al. (Ref 58). A higher amount of precipitation may affect the separation of grain boundaries, which reduces the tensile properties.

3.4 Residual Stresses

During solidification of welded material, the top and bottom part of the fusion region cools rapidly compared to HAZ. Uneven temperature distribution occurs in the weldment owing to rapid heating and consecutive cooling, resulting in the development of residual stresses. The evolution of residual stress in welded joints is mainly controlled by the metallurgical phase transformations and related microstructure, which impact the weld strength properties (Ref 41). During welding of thin steel plates, residual stresses may account for the development

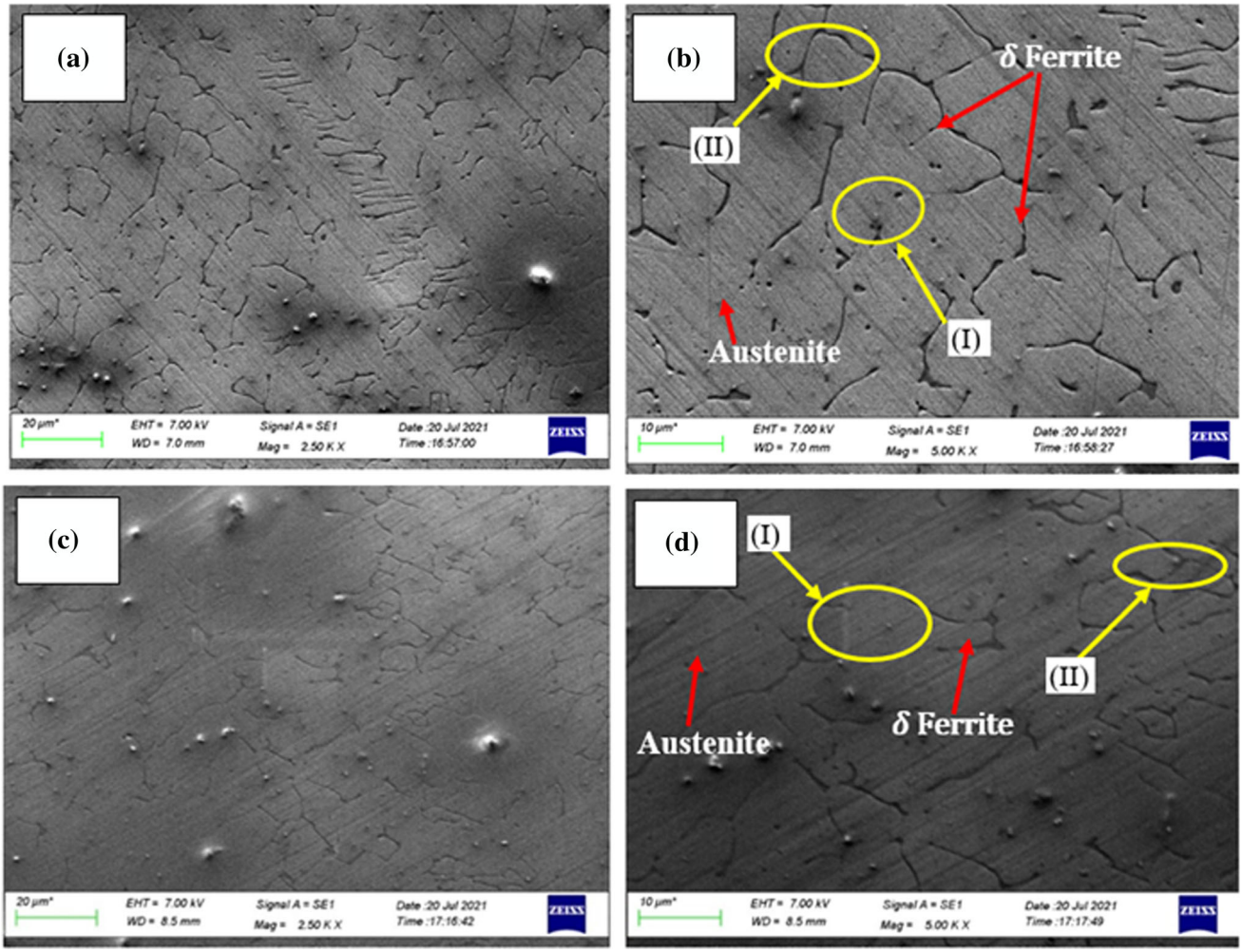


Fig. 7 SEM images of sample 9 (CMT) (a) and sample 6 (U-CMT) (c) and (b) and (d) are enlarged views

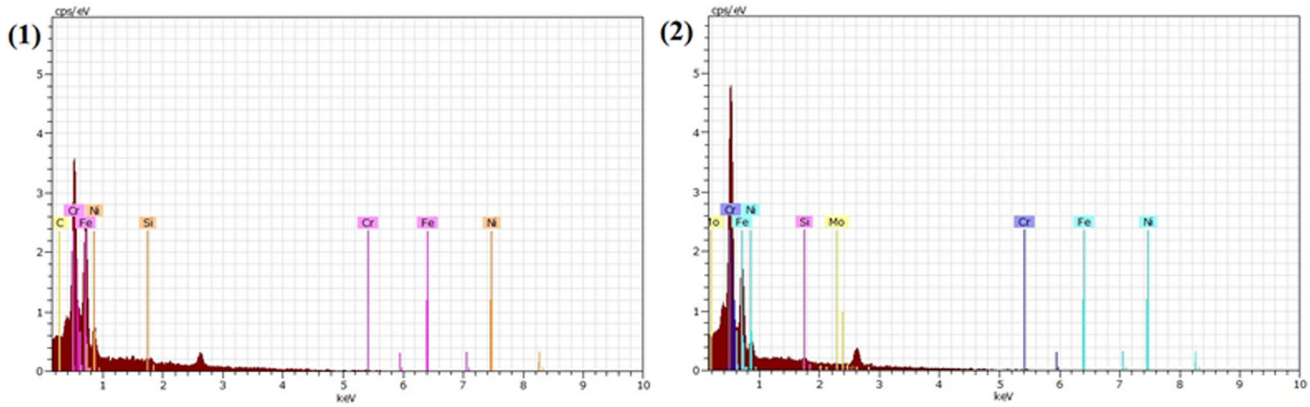


Fig. 8 EDS analysis of sample 6 (U-CMT) taken at different regions in SEM micrograph

of distortion or buckling type welding defects (Ref 59). By the application vibrations during welding, the generation of residual stress may be efficiently reduced to a large extent. The residual stress measurements are taken at the cross section of welded samples at many regions, and the results are displayed in Fig. 14. The fusion region has higher residual stress than other regions. U-CMT (95-4-7) sample 6 welded with 99 μm amplitude has lower residual stresses than other welded samples. Figure 15 exhibits the Debye–Scherrer ring, distortion, and residual profile curve of welded samples. In the Debye-Scherrer ring, the red color shows more residual stresses in the tip, and the blue color-shaded bottom indicates a minor concentration of residual stresses. A higher number of distortions was observed in the CMT sample due to the larger value of residual stresses, as seen in Fig. 15(b). Aoki et al. (Ref 60), Qinghua et al. (Ref 61), and Chuvas et al. (Ref 62) observed that residual stresses are reduced in the fusion region with vibration in comparison with the non-vibration welded joint. With the increase in vibrational amplitude, a higher number of residual stresses will be minimized compared to lower vibrational amplitude welded joints (Ref 56). Hsieh et al. (Ref 55) reported that there is no notable development in residual stress relief of TIG 304 SS welded joints with higher vibrational frequency. U-CMT sample has lower residual stresses with higher tensile strength due to grain refinement occurring with vibration during welding, resulting in improved joint efficiency of the welded joint. The highest tensile strength obtained sample is taken to check the weld quality by NDT radiography inspection method as per ASTM SectionV: 2017 standard. Figure 16 distinctly shows that the welded joint does not consist of any welding defects such as incomplete penetration, cracks, incomplete fusion, and porosity. It reveals that the fabricated welded joint has a sound welded joint with quality.

Table 3 EDS main elements in the SEM micrograph (Fig. 7) of CMT and U-CMT (in w%)

	Region	Cr	Ni	Fe
CMT	I	17.24	9.24	69.87
	II	21.46	7.68	70.88
U-CMT	I	19.77	10.86	65.94
	II	23.84	8.52	64.16

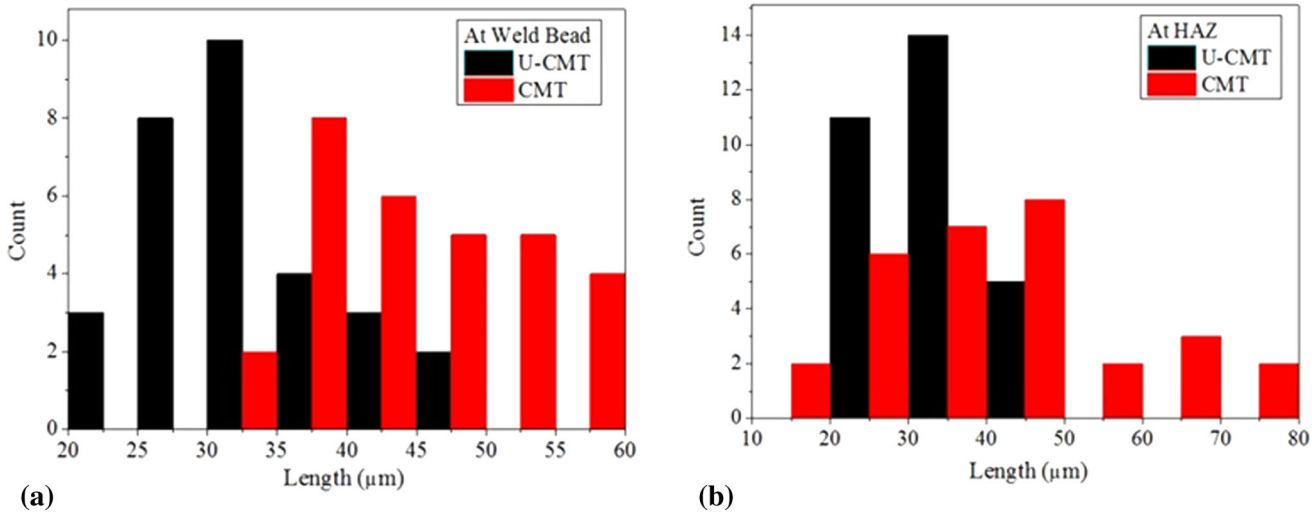


Fig. 9 Grain size comparison for U-CMT and CMT samples at (a) Fusion region (b) HAZ

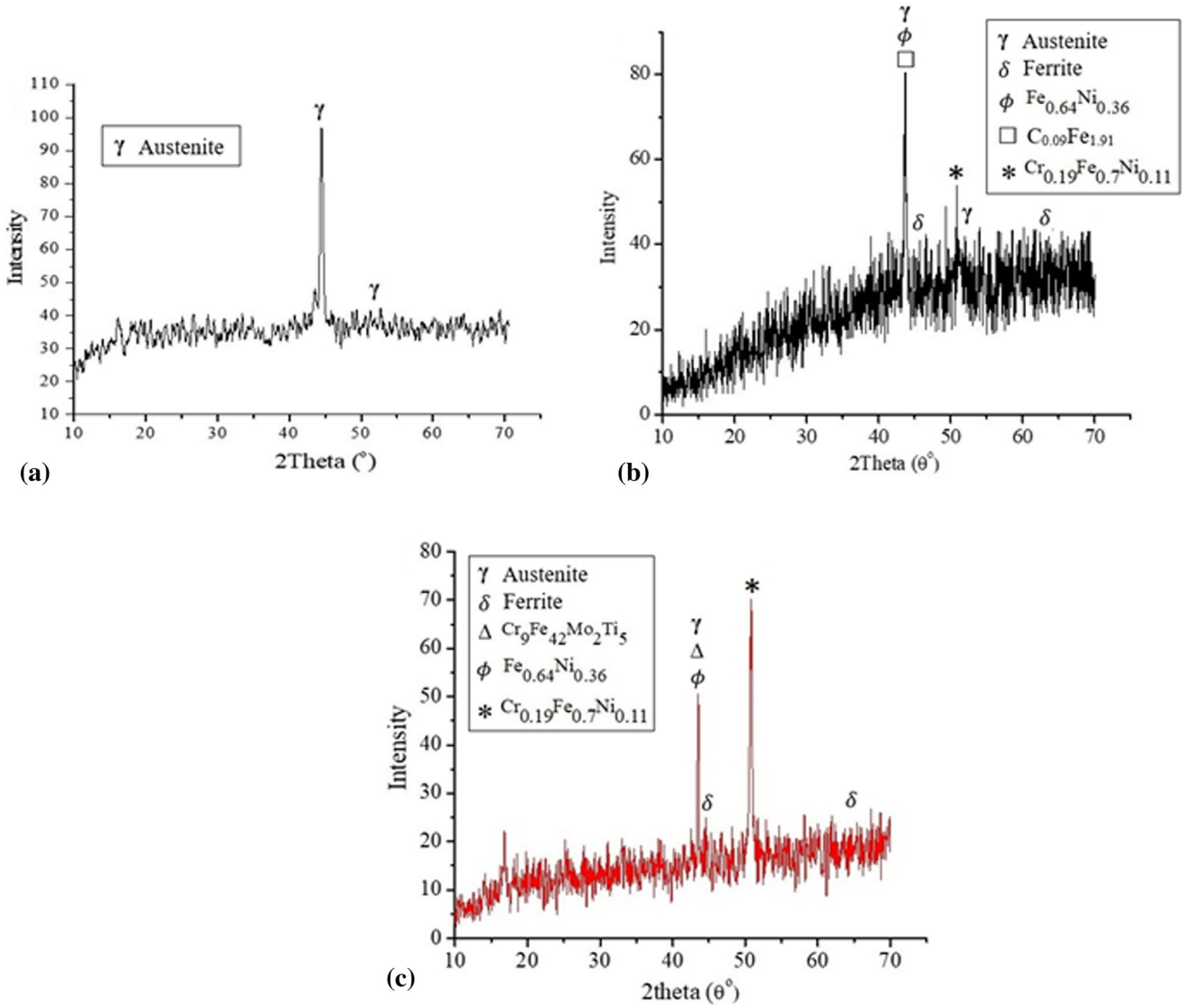


Fig. 10 XRD plots of (a) Base material (Ref 53) (b) CMT and (c) U-CMT welded samples

4. Conclusions

This work aims to investigate the effect of ultrasonic vibration on CMT welding of AISI 304 stainless steel joints. Based on experimental results, the main conclusions are summarized as follows:

1. Good weld strength is exhibited in all U-CMT samples with various welding parameters, which shows that ultrasonic vibration has a significant role in enhancing weld joint strength.
2. The microstructure of the fusion region consists of austenite and δ-ferrite phases. With the application of vibrations, the equiaxed refined grains are formed.
3. The microhardness of U-CMT samples is approximately 16% increase in comparison with CMT welded samples, and higher tensile strength of 725 MPa is attained with the utilization of vibration of the welded joint.

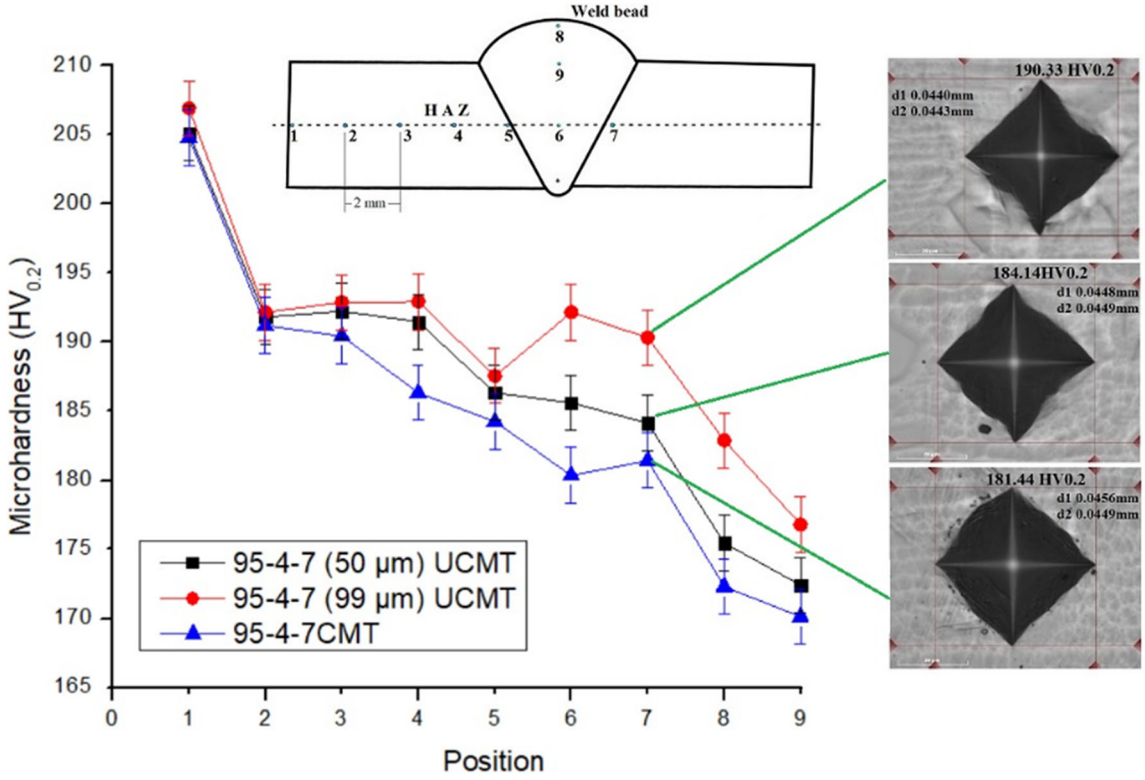


Fig. 11 Microhardness variation of welded joints

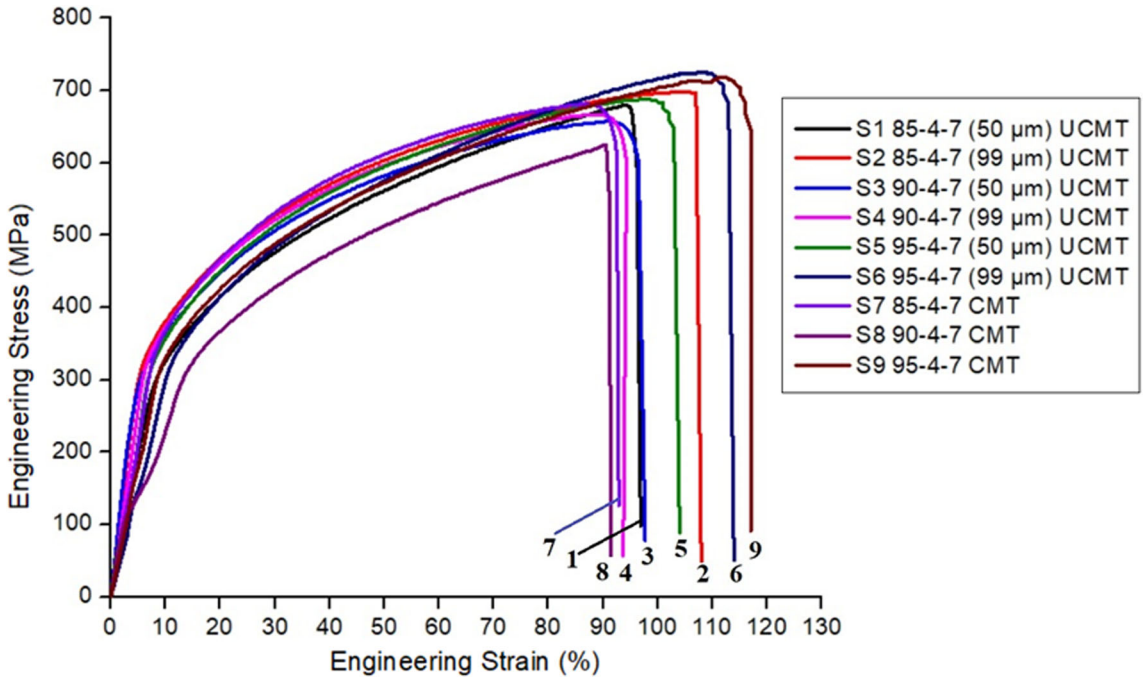


Fig. 12 Stress–strain curves of welded samples

4. Tensile fracture surface reveals fine dimpled morphology with micro voids and sub-microvoids observed in U-CMT welded samples.
5. Significant reduction of residual stresses detected with vibration treatment during welding, which enhances joint efficiency without any defects.

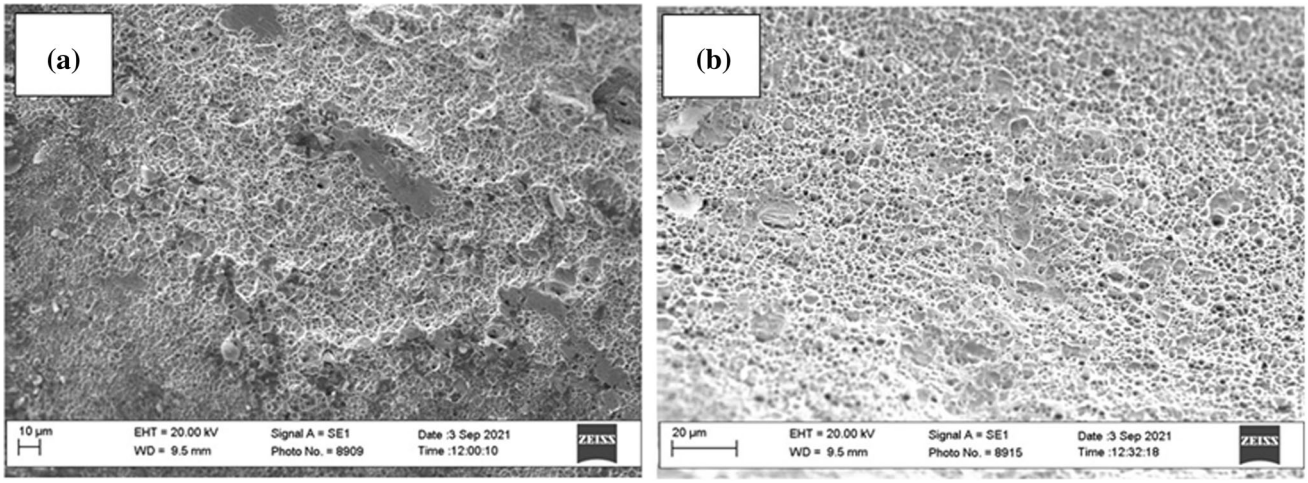


Fig. 13. FESEM fractography of tensile fractured surface of (a) CMT sample -9 and (b) U-CMT sample-6

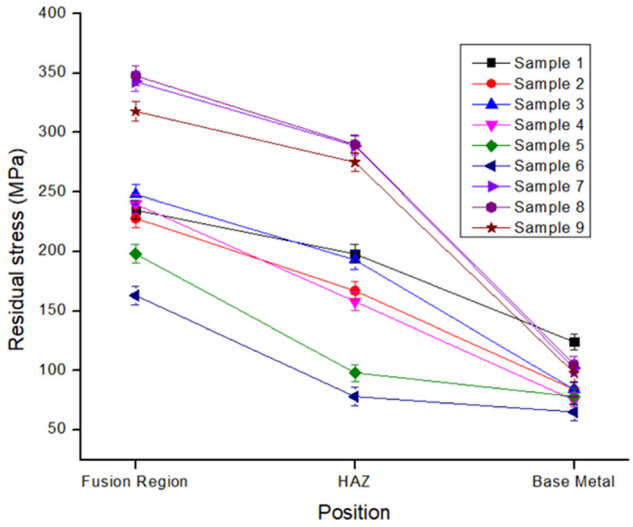


Fig. 14 Residual Test results of welded samples taken at various points

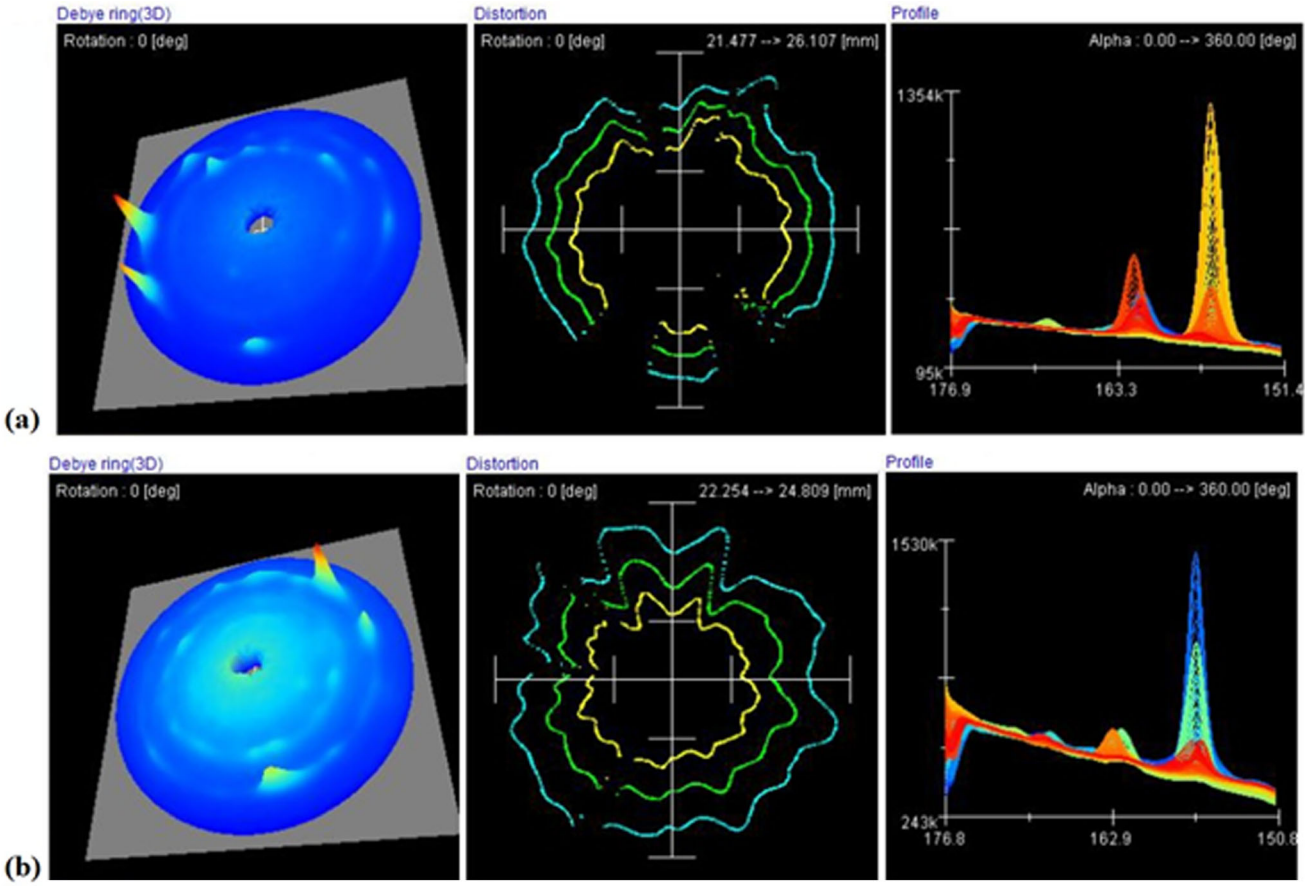


Fig. 15 Residual stress distortion and profile curve of (a) Lower residual stress sample 6 (U-CMT) and (b) Higher residual stress sample 8 (CMT) at Fusion region

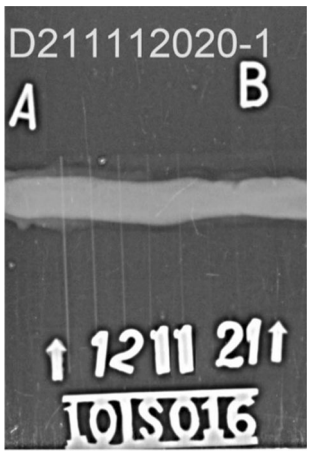


Fig. 16 Radiography inspection test of U-CMT sample 6

References

1. N.R. Baddoo, Stainless Steel in Construction: A Review of Research, Applications, Challenges and Opportunities, *J Constr. Steel Res.*, 2008, **64**, p 1199–1206
2. S. Saha, M. Mukherjee, and T.K. Pal, Microstructure, Texture, and Mechanical Property Analysis of Gas Metal Arc Welded AISI 304 Austenitic Stainless Steel, *J. Mater. Eng. Perform.*, 2015, **24**, p 1125–1139
3. P.V.S.S. Sridhar, P. Biswas, and P. Mahanta, Effect of Process Parameters on Bead Geometry, Tensile and Microstructural Properties of Double-Sided Butt Submerged Arc Welding of SS 304 Austenitic Stainless Steel, *J. Braz. Soc. Mech. Sci. Eng.*, 2020, **42**, p 551
4. E.O. Ogundimu, E.T. Akinlabi, and M.F. Erinosho, Comparative Study Between TIG and MIG Welding Processes, *Int. Conf. Eng. Sust. World, J. Phys. Conf. Ser.*, 2019, **1378**, p 022074
5. S. Selvi, A. Vishvaksenan, and E. Rajasekar, Cold Metal Transfer Technology-An Overview, *Def. Technol.*, 2021, **14**, p 28–44
6. Y. Koli, N. Yuvaraj, S. Aravindan, and Vipin, CMT Joining of AA6061-T6 and AA6082-T6 and Examining Mechanical Properties and Microstructural Characterization, *Trans. Indian Inst. Met.*, 2021, **74**, p 313–329
7. S. Zhou, G. Ma, D. Wu, D. Chai, and M. Lei, Ultrasonic Vibration Assisted Laser Welding of Nickel-Based Alloy and Austenite Stainless Steel, *J. Manuf. Process.*, 2018, **31**, p 759–767
8. Y. Wang, C. Yu, H. Lu, and J. Chen, Research Status and Future Perspectives on Ultrasonic Arc Welding Technique, *J. Manuf. Process.*, 2020, **58**, p 936–954
9. M.J. Jose, S.S. Kumar, and A. Sharma, Vibration Assisted Welding Processes and their Influence on Quality of Welds, *Sci. Technol. Weld. Join.*, 2016, **21**, p 243–258
10. S. Kumar, C.S. Wu, G.K. Padhy, and W. Ding, Application of Ultrasonic Vibrations in Welding and Metal Processing: A Status Review, *J. Manuf. Process.*, 2017, **26**, p 295–322
11. F. Ning and W. Cong, Ultrasonic Vibration-Assisted (UV-A) Manufacturing Processes: State of the Art and Future Perspectives, *J. Manuf. Process.*, 2020, **51**, p 174–190
12. T.V. Cunha and C.E.N. Bohorquez, Ultrasound in Arc Welding: A Review, *Ultrasonics*, 2015, **56**, p 201–209
13. T. Watanabe, M. Shiroki, A. Yanagisawa, and T. Sasaki, Improvement of Mechanical Properties of Ferritic Stainless Steel Weld Metal by Ultrasonic Vibration, *J. Mater. Process. Technol.*, 2010, **210**, p 1646–1651

14. W. Xie, T. Huang, C. Yang, C. Fan, S. Lin, and W. Xu, Comparison of Microstructure, Mechanical Properties, and Corrosion Behavior of Gas Metal Arc (GMA) and Ultrasonic-Wave-Assisted GMA (U-GMA) Welded Joints of Al–Zn–Mg Alloy, *J. Mater. Process. Technol.*, 2020, **277**, p 116470
15. W.L. Dai, Effects of High-Intensity Ultrasonic-Wave Emission on the Weldability of Aluminum Alloy 7075–T6, *Mater. Lett.*, 2003, **57**, p 2447–2454
16. G. Zhao, Z. Wang, S. Hu, S. Duan, and Y. Chen, Effect of Ultrasonic Vibration of Molten Pool on Microstructure and Mechanical Properties of Ti-6Al-4V Joints Prepared via CMT+ P Welding, *J. Manuf. Process.*, 2020, **52**, p 193–202
17. Q.H. Chen, S.B. Lin, C.L. Yang, C.L. Fan, and H.L. Ge, Effect of Ultrasound on Heterogeneous Nucleation in TIG Welding of Al–Li Alloy, *Acta. Metall. Sin. Eng.*, 2016, **29**, p 1081–1088
18. T. Yuan, S. Kou, and Z. Luo, Grain Refining by Ultrasonic Stirring of the Weld Pool, *Acta. Mater.*, 2016, **106**, p 144–154
19. Y. Tian, J. Shen, S. Hu, Z. Wang, and J. Gou, Effects of Ultrasonic Vibration in the CMT Process on Welded Joints of Al Alloy, *J. Mater. Process. Tech.*, 2018, **259**, p 282–291
20. M. Fattahi, A. Ghaheri, N. Arabian, F. Amirkhanlu, and H. Moayedi, Applying the Ultrasonic Vibration During TIG Welding as a Promising Approach for the Development of Nanoparticle Dispersion Strengthened Aluminum Weldments, *J. Mater. Process. Technol.*, 2020, **282**, p 116672
21. F. Ning and W. Cong, Microstructures and Mechanical Properties of Fe-Cr Stainless Steel Parts Fabricated by Ultrasonic Vibration-Assisted Laser Engineered Net Shaping Process, *Mater. Lett.*, 2016, **179**, p 61–64
22. C. Chen, C. Fan, X. Cai, S. Lin, C. Yang, and Y. Zhuo, Microstructure and Mechanical Properties of Q235 Steel Welded Joint in Pulsed and Un-Pulsed Ultrasonic Assisted Gas Tungsten Arc Welding, *J. Mater. Process. Technol.*, 2020, **275**, p 116335
23. C. Chen, C. Fan, S. Lin, X. Cai, L. Zhou, S. Ye, and C. Yang, Effect of Ultrasonic Pattern on Weld Appearance and Droplet Transfer in Ultrasonic Assisted MIG Welding Process, *J. Manuf. Process.*, 2018, **35**, p 368–372
24. C. Chen, C. Fan, X. Cai, S. Lin, and C. Yang, Analysis of Droplet Transfer, Weld Formation and Microstructure in Al-Cu Alloy Bead Welding Joint with Pulsed Ultrasonic-GMAW Method, *J. Mater. Process. Tech.*, 2019, **271**, p 144–151
25. F. Fan, L. Zhou, Z. Liu, C. Yang, S. Lin, W. Xie, and H. Tong, Arc Character and Droplet Transfer of Pulsed Ultrasonic Wave-Assisted GMAW, *Int. J. Adv. Manuf. Technol.*, 2018, **95**, p 2219–2226
26. C. Hua, H. Lua, C. Yu, J. Chen, X. Wei, and J. Xu, Reduction of Ductility-Dip Cracking Susceptibility by Ultrasonic-Assisted GTAW, *J. Mater. Process. Technol.*, 2017, **239**, p 240–250
27. H. Lan, X. Gong, S. Zhang, L. Wang, B. Wang, and L.-P. Nie, Ultrasonic Vibration Assisted Tungsten Inert Gas Welding of Dissimilar Metals 316L and L415, *Int. J. Miner. Metall. Mater.*, 2020, **27**, p 943–953
28. Y. Li, S. Tian, C. Wu, and M. Tanaka, Experimental Sensing of Molten Flow Velocity, Weld Pool and Keyhole Geometries in Ultrasonic-Assisted Plasma Arc Welding, *J. Manuf. Process.*, 2021, **64**, p 1412–1419
29. A.V. Kolubaev, O.V. Sizova, S.V. Fortuna, A.V. Vorontsov, A.N. Ivanov and E.A. Kolubaev, Weld Structure of Low-Carbon Structural Steel Formed by Ultrasonic-Assisted Laser Welding, *J. Constr. Steel Res.*, 2020, **172**, p 106190
30. S. Yu, A.V. Tarasov, S.V. Vorontsov, V.E. Rubtsov, V.A. Krasnoveikin, and E.A. Kolubaev, Ultrasonic-Assisted Laser Welding on AISI 321 Stainless Steel, *Weld. World.*, 2019, **63**, p 875–886
31. A.I. Gorunov, O.A. Nyukhlaev, and A. Kh, Investigation of Microstructure and Properties of Low-Carbon Steel During Ultrasonic-Assisted Laser Welding and Cladding, *Int. J Adv. Manuf. Technol.*, 2018, **99**, p 2467–2479
32. Y. Ou, O. Lu, C. Li, H. Yan, P. Zhang, and J. Jin, Effect of Vibration Frequency on Microstructure and Properties of Laser-Welded Inconel 718 Nickel-Base Superalloy, *J. Mater. Eng. Perform.*, 2021, **30**, p 2399–2407
33. J. Jin, Q. Lu, P. Zhang, C. Li, and H. Yan, Research on Microstructure and Fatigue Properties of Vibration-Assisted 5052 Aluminum Alloy Laser Welded Joints, *J. Mater. Eng. Perform.*, 2020, **29**, p 4197–4205
34. H. Chen, N. Guo, Z. Zhang, C. Liu, L. Zhou and G. Wang, A Novel Strategy for Metal Transfer Controlling in Underwater Wet Welding Using Ultrasonic-Assisted Method, *Mater. Lett.*, 2020, **270**, p 127692
35. N.S. Rossini, M. Dassisti, K.Y. Benyounis, and A.G. Olabi, Methods of Measuring Residual Stresses in Components, *Mater. Des.*, 2012, **35**, p 572–588
36. P.K. Taraphdar, R. Kumar, C. Pandey, and M.M. Mahapatra, Significance of Finite Element Models and Solid-State Phase Transformation on the Evaluation of weld Induced Residual Stresses, *Met. Mater. Int.*, 2021 https://doi.org/10.1007/s12540-020-00921-4
37. C. Pandey, M.M. Mahapatra, P. Kumar, and N. Saini, Effect of Weld Consumable Conditioning on the Diffusible Hydrogen and Subsequent Residual Stress and Flexural Strength of Multipass Welded P91 Steels, *Metall. Mater. Trans. B.*, 2018, **49B**, p 2881–2895
38. S.D. Banik, S. Kumar, and P.K. Singh, Distortion and Residual Stresses in Thick Plate Weld Joint of Austenitic Stainless Steel: Experiments and Analysis, *J Mater. Process. Technol.*, 2021, **289**, p 116944
39. P.K. Taraphdar, J.G. Thakare, C. Pandey, and M.M. Mahapatra, Novel Residual Stress Measurement Technique to Evaluate Through Thickness Residual Stress Fields, *Mater. Lett.*, 2020, **277**, p 128347
40. H. Lai and W. Wu, Practical Examination of the Welding Residual Stress in View of Low-Carbon Steel Welds, *J Mater. Res. Technol.*, 2020, **9**, p 2717–2726
41. B. Kumar, S. Bag, S. Mahadevan, C.P. Paul, C.R. Das, and K.S. Bindra, On the Interaction of Microstructural Morphology with Residual Stress in Fiber Laser Welding of Austenitic Stainless Steel, *CIRP J. Manuf. Sci. Technol.*, 2021, **33**, p 158–175
42. R. Gadallah, S. Tsutsumi, Y. Aoki, and H. Fujii, Investigation of Residual Stress Within Linear Friction Welded Steel Sheets by Alternating Pressure via X-ray Diffraction and Contour Method Approaches, *J. Manuf. Process.*, 2021, **64**, p 1223–1234
43. D.A. Lesyk, H. Soyama, B.N. Mordyuk, V.V. Dzhemelinskyi, S. Martinez, N.I. Khripta, and A. Lamikiz, Mechanical Surface Treatments of AISI 304 Stainless Steel: Effects on Surface Microrelief, Residual Stress and Microstructure, *J. Mater. Eng. Perform.*, 2019, **28**, p 5307–5322
44. M. Vykunta Rao, P. Srnivasa Rao, and B. Surendra Babu, Vibratory Weld Conditioning During Gas Tungsten Arc Welding of Al 5052 Alloy on the Mechanical and Micro-Structural Behavior, *World J. Eng.*, 2020, **17**(6), p 831–836
45. J. Singh and A.S. Shahi, Weld Joint Design and Thermal Aging Influence on the Metallurgical, Sensitization and Pitting Corrosion Behavior of AISI 304L Stainless Steel Welds, *J. Manuf. Process.*, 2018, **33**, p 126–135
46. Q.J. Sun, S.B. Lin, C.L. Yang, and G.Q. Zhao, Penetration Increase of AISI 304 Using Ultrasonic Assisted Tungsten Inert Gas Welding, *Sci. Technol. Weld. Join.*, 2009, **14**, p 765–767
47. C.-W. Kuo, S.-M. Yang, J.-H. Chen, G.-H. Lai and W. Wu, Study of Vibration Welding Mechanism, *Sci. Technol. Weld. Join.*, 2008, **13**, p 357–362
48. B. Wei, Unidirectional Dendritic Solidification Under Longitudinal Resonant Vibration, *Acta Metall. Mater.*, 1992, **40**, p 2739–2751
49. Q. Chen, S. Lin, C. Yang, C. Fan, and H. Ge, Grain Fragmentation in Ultrasonic-Assisted TIG Weld of Pure Aluminum, *Ultrason. Sonochem.*, 2017, **39**, p 403–413
50. C. Pandey, Mechanical and Metallurgical Characterization of Dissimilar P92/SS304 L Welded Joints Under Varying Heat Treatment Regimes, *Metall. Mater. Trans. A.*, 2020, **51A**, p 2126–2142
51. G.R. Mirshekari, E. Tavakoli, M. Atapour, and B. Sadeghian, Microstructure and Corrosion Behavior of Multipass Gas Tungsten Arc Welded 304L Stainless Steel, *Mater. Des.*, 2014, **55**, p 905–911
52. S. Balbande, R.V. Taiwade, A.P. Patil, and U. Pathak, Cold Metal Transfer Welding of Duplex with Ferritic and Super Austenitic Stainless Steel, *Mater. Res. Express*, 2019, **6**, p 106542
53. J.G. Roy, N. Yuvaraj, and Vipin, Effect of Welding Parameters on Mechanical Properties of Cold Metal Transfer Welded Thin AISI 304 Stainless-Steel Sheets, *Trans. Indian Inst. Met.*, 2021, **74**, p 2397–2408
54. S. Kumar and A.S. Shahi, Effect of Heat Input on the Microstructure and Mechanical Properties of Gas Tungsten Arc Welded AISI 304 Stainless Steel Joints, *Mater. Des.*, 2011, **32**, p 3617–3623
55. C. Hsieh, P. Wang, J. Wang, and W. Wu, Evolution of Microstructure and Residual Stress Under Various Vibration Modes in 304 Stainless Steel Welds, *Sci. World J.*, 2014 https://doi.org/10.1155/2014/895790

56. C. Hsieh, C. Lai, and W. Wu, Effect of Vibration on Microstructures and Mechanical Properties of 304 Stainless Steel GTA Welds, *Metals. Mater. Int.*, 2013, **19**, p 835–844
57. Y. Koli, N. Yuvaraj, S. Vipin, and S. Aravindan, Investigations on Weld Bead Geometry and Microstructure in CMT, MIG Pulse Synergic and MIG Welding of AA6061-T6, *Mater. Res. Express*, 2020, **6**, p 1265e5
58. J. Yan, M. Gao and X. Zeng, Study on Microstructure and Mechanical Properties of 304 Stainless Steel Joints by TIG, Laser and Laser-TIG Hybrid Welding, *Opt. Lasers Eng.*, 2010, **48**, p 512–517
59. H. Eisazadeh and D.K. Aidun, Residual Stress Reduction in Dissimilar Metals Weld, *J. Manuf. Process.*, 2021, **64**, p 1462–1475
60. S. Aoki, T. Nishimura and T. Hiroi, Reduction Method for Residual Stress of Welded Joint Using Random Vibration, *Nucl. Eng. Des.*, 2005, **235**, p 1441–1445
61. L. Qinghua, C. Ligong, and N. Chunzhen, Effect of Vibratory Weld Conditioning on Welded Valve Properties, *Mech. Mater.*, 2008, **40**, p 565–574
62. T.C. Chuvas, D.A. Castello and M.P.C. Fonseca, Residual Stress Relief of Welded Joints by Mechanical Vibrations, *J. Braz. Soc. Mech. Sci. Eng.*, 2016, **38**, p 2449–2457

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

International Journal of Thermodynamics (IJoT)
ISSN 1301-9724 / e-ISSN 2146-1511
https://dergipark.org.tr/en/pub/ijot

Vol. xx (No. x), pp. xxx-xxx, 2022
doi: 10.5541/ijot.PID
Published online: Mmm dd, 2022

Enhancing the Thermal Performance of a Double Pipe Heat Exchanger in Turbulent Flow Conditions

Manish Sanserwal[1,2], Devendra Yadav[1*], and Mayank Bhardwaj[3], Gurjeet Singh[4]

[1]Department of Mechanical Engineering, Galgotias College of Engineering and Technology, Greater Noida 201308, Uttar Pradesh, India
[2]Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[3]University Institute of Engineering and Technology, Kurukshetra University, Kurukshetra, Haryana, India
[4]Department of Mechanical Engineering, PEC University of Technology, Chandigarh, India
E-mail: [1*]ydevendra393@gmail.com

Received 4 February 2022, Revised 13 April 2022, Accepted 27 April 2022

Abstract

Heat exchangers with high thermal performance are required for industrial applications. Using heat transfer methodology in conjunction with simple design changes and assembly functions of heat exchangers could be an effective way to accomplish this. An experimental analysis was performed in this study to improve the heat transfer performance of a double pipe heat exchanger by implanting a flat strip spring turbulator (FST) within the heat exchanger's inner tube. The experimental investigation of the Double pipe heat exchanger in conjunction with three sets of FST turbulators (pitch: 15 cm, 10 cm, and 5 cm) for turbulent flow (Re 9000-38000) was carried out. The Nusselt number, friction factor ratio, and thermal performance factor of heat exchangers with FST at various pitches are found to be between 60 and 170, 1.44 and 1.76, and 0.94 and 1.06, respectively. The highest heat transfer achieved by using a flat spring turbulator is 20% for a pitch value of 5cm. In comparison to other sets of FST, a double pipe heat exchanger with FST pitch value of 10 cm has greater thermohydraulic performance. When compared to previous research, the experimental results obtained from this work at higher Reynolds numbers the friction factor are within a well-accepted range.

Keywords: *Heat exchanger; spring turbulator; thermal performance factor; heat transfer coefficient; Wilson plot.*

1. Introduction

The heat exchanger allows heat to be transferred from the hotter fluid to the cooler fluid. There are essentially two types of categories: direct and indirect. In comparison to direct heat exchangers, indirect heat exchangers are more commonly used in industries since they eliminate fluid mixing during operation. In the current context, designing a heat exchanger is not an easy task because it still has scale and fluid flow rate constraints depending on the application. To obtain high heat transferability in heat exchangers, more attention on size reduction is required. The numerous techniques used in the heat exchanger to improve the rate of heat transfer [1]–[4], can be categorized as active, passive, and combined techniques (Fig. 1). Mechanical aids, injection, suction, electrostatic fields, and surface and fluid vibration all demand more power (external power) than the power used to run the heat exchanger. In contrast, in passive approaches, specially engineered geometries or turbulent circulation generators or turbulators are employed to impede the fluid flow with the purpose of enhancing heat transfer without the use of an external (additional) power source [5]. Multiple heat transfer improvement approaches, such as the use of twisted strips and tapes, coil or helical wire, polished surfaces, rough surfaces, stretched surfaces, perforated conical rings, conical springs, and so on, are included in the passive approach. However, the compound strategy combines passive and active methods to improve the thermo-hydraulic performance of a heat exchanger. Hence, heat transfer enhancement in a heat exchanger can also be procured by creating turbulence in the fluid flow and at last, considering this concept as a motivation for literature. Therefore, the literature study was carried out to investigate how much higher the heat transfer rate in the heat exchanger with turbulators can be achieved. For the improvement in heat transfer rate with a full width twisted tape under laminar flow and steady wall temperature condition, Dasmahapatra & Rao [6] utilizes a viscous non-Newtonian fluid. Al-Fahed & Chakroun [7] experimentally investigated the heat transfer enhancement in a fully developed turbulent flow with a tube-tape clearance under constant heat flux condition. Whereas, an experimental study on twisted tape turbulator in a horizontal tube under viscous flow conditions were carried out by Manglik & Bergles [8]. Zamankhan [9] studied an improvement in the heat transfer rate in a heat exchanger with a helical metal wire turbulator using a glycol-water solution as a working fluid with varying concentrations. The 3D mathematical model, also developed for the confirmation of experimental data and the comparison of numerical findings with experimental results, concluded that actual system behaviors could be predicted by the LES model. [10] investigate conical spring turbulators in different configurations (convergent, divergent, and convergent-divergent conical rings CR, DR, and CDR) at different cone angles of 30°, 45°, 60° in a concentric double pipe heat

*Corresponding Author

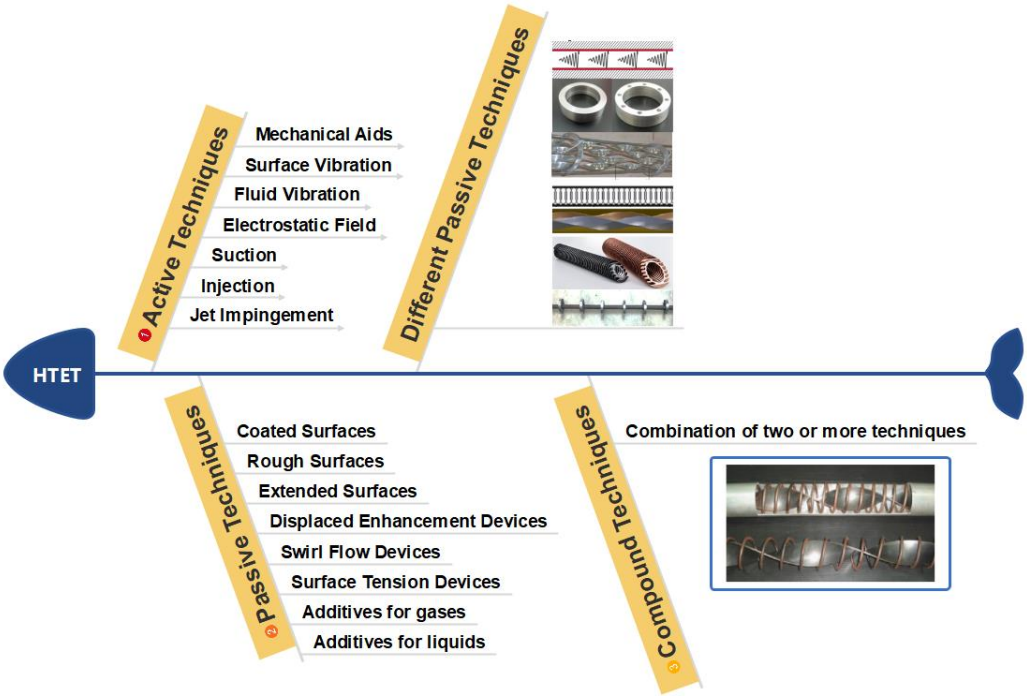


Figure 1. Techniques for heat transfer enhancement (HTET).

exchanger at different Reynolds number (10000-34000). Yadav et al. [11], [12] designed and fabricated a trio tube heat exchanger setup which has the better heat transfer capability and also compact in size. This heat exchanger requires ≈ 58% smaller in pipe length for the same amount of heat transfer as of double pipe heat exchanger. Sheikholeslami et al. [13]–[15] conducted an experimental investigation on a double pipe air to water heat exchanger with discontinuous helical turbulators (typical plane and perforated) at different Reynolds numbers, pitch and open area ratio for estimating the behaviour of heat transfer and pressure drop. For finding the optimal design of heat exchanger, Non-dominated Sorting Genetic Algorithm II (NSGA II) is used for having high efficiency and ANSYS FLUENT14 for better numerical simulation.

Later on, in the same scenario, the investigation was conducted typical and perforated circular-ring (TCR and PCR) turbulators. Nanan et.al. [16] carried out a comparative investigation in a heat exchanger between different turbulators designs: twisted and straight cross-baffles, twisted-baffles, alternate twisted and straight alternate-baffles and last one is straight baffles and with different pitch ratios (P/D = 1 to 2) and Reynolds number (6000 to 20000). For better comparison, a numerical simulation also is done with all types of turbulator for a better understanding of heat transfer enhancement and friction factor. Mashoofi et. al. [17] investigated tube in tube helically coil (TTHC) heat exchange with and without helical wire turbulator in four ways: TTHC heat exchanger a) with turbulator inside the inner tube b) with turbulator inside the annulus c) with turbulator inside both tube d) without turbulator, for evaluating the effect on heat transfer and frictional factor. The use of turbulator only in the annulus (containing hot water) and turbulator only in an inner tube (containing air) enhance the airside Nusselt number by 8-32% and 52-82%, respectively. Later, a helical wire turbulator (only inside the tube) in the shell and tube helically coiled heat exchanger was investigated by Panahi et al. [18]. Sandeep et al. [19] experimentally and numerically investigated a novel turbulator (aluminum small plate placed in the cross-type arrangement) act as airflow divider at a different pitch to tube diameter ratios varying from 0.54 to 1.09 at a 90° angle of twist. For evaluating the Nusselt number enhancement at a different angle of twist (45° and 30°), a CFD simulation was conducted and find out 1.33 to 1.46 times and 1.43 to 1.60 times of enhancement at 45° and 30°, respectively. Khorasani et al. [20] investigated the effect of a spiral wire turbulator with four different-different spring pitches and wire diameter in a helical tube with constant heat flux. Further, each arrangement was conducted for five types of flow rates of water. It is found that, with the increase in spring pitch and wire diameter of the spiral wire turbulator, Nusselt number also increases up to 70% and 73% respectively. Zohiret. al. [21] utilize a coiled wire turbulator upon the outer surface of the inner tube of double pipe heat exchanger and achieve convective heat transfer coefficient enhancement of 400% and 450% in parallel (same direction) and counterflow (opposite direction) respectively. Budaket. al. [22] numerically analyzes the four geometries of turbulators in concentric pipe heat exchanger located inside the inner pipe and considering both parallel flow and counter flow condition at different flow rates. Also, formed an ANSYS 12.0 fluent program code to analyses pressure and thermal characteristics. Kumar et. al. [23] included the effect of perforation index (PI = 8% to 24%) and found, 4 and 1.47 times of heat transfer enhancement at PI=8% & d/D=0.6 and PI=24% & d/D=0.8 condition, respectively when compared with the plain tube. Singh et. al. [24] experimentally investigated circular solid ring turbulator with multiple twisted tape arrangements inside the core. Later, Kumar et al.[25] utilized both solid and perforated circular-ring turbulator with twisted tape for investigation. Results revealed improvement in both, heat transfer and thermal performance factor over the smooth pipe in a range of around 2.2-3.54 and 1.18- 1.64 times, respectively. Whereas, Dattet. al. [26] investigated a solid circular ring turbulator with a number (ranges 1 to 4) of square wing twisted tape. Akpinar [27] experimentally studied the effect of helical spring turbulator inside the inner pipe of a double pipe heat exchanger on heat transfer and friction factor. Nusselt

number and dimensionless exergy loss increment found to be 2.64 and 1.16 times, respectively as compared to the heat exchanger not using turbulator. Maradiya et. al. [28] revealed that twisted tape as turbulator not performed well with air as compare to water as a working fluid due to large density of liquid. Also, in case of air heating application, ribs or deflector and vortex generators, whereas, in case of liquids, swirl producing devices are more useful in thermal performance factor improvement.

According to the literature and recent review papers [29]–[31], passive turbulators perform better in the water medium than in the air medium in the double concentric pipe heat exchanger. In most cases, turbulators clearly increase the heat transfer rate to a significant level at a high Reynolds number. The majority of the researches concentrated on disturbing the fluid not only in the centre but also along the wall of the heat exchanger tube (where turbulators are inserted) in order to disturb the laminar sub-layer. The current study used a flat strip spring turbulator (FST) to alleviate the disadvantages associated with earlier investigations, namely the higher value of the friction factor. This FST design advantage of less material use and to expect the maximum possible increase in heat transfer at the lowest pressure drop. So far, no experimental work on flat metal strip springs has been published; this is a novel design consisting of circular rings, springs, and twisted tape to provide better distribution of fluid streams with lower frictional loss. This research also focuses on the employment of several sets of FST turbulator to obtain the best FST value for maximum thermohydraulic performance.

2. Experimental

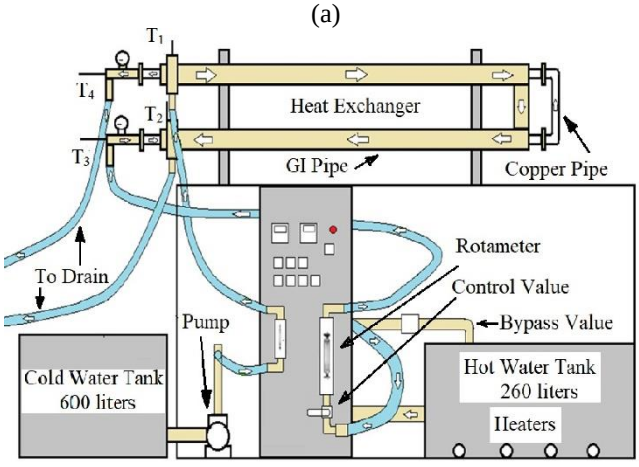
2.1 Fabrication of Experimental Setup

The theoretical analysis of the double pipe heat exchanger helped to develop the final size of the heat exchanger. Well before its mountings and attachments, the new apparatus underwent primary calibration. The construction of the experimental setup is made of mild steel to make the apparatus stable and sturdy. The experimental setup consists of housing for hot water tanks, cold and hot water pumps, display boards and test pipes. Two cold water tanks were used, one for the supply of fresh low-temperature cold water and the other for the storage of high-temperature cold water from the test section. Coldwater (at room temperature) and hot water (at set point 76 ±1℃) were allowed to flow through the annulus and inner pipe section, respectively, in the counter-flow direction. A schematic diagram of the experimental setup is shown in Fig. 2.

The test section consists of smooth inner copper tubes of 4432 mm length and outer (OD) and inner (ID) diameters of 24.5 mm and 21.5 mm respectively. The tubes were brought in, three separate parts of 2m+2m and a bend of 0.232 m. The outer G.I. pipes were selected accordingly so that there remains an effective inside diameter sufficient enough to maintain adequate flow and not to alter the original flow rate having an outer and inner diameter of 45mm and 50 mm respectively. The U-bend section is made detachable by providing flange couplings on both ends of individual lengths of copper tubes. This detachable portion was utilized to insert a full-length FST setup (when mounted on rods) easily. Two rotameters (glass tube flow meter) i.e. a cold water side small and a hot water side large were employed, with a maximum range of 100 to 1500 and 100 to 2000 LPH (liters per hour) respectively. The cold water (at room temperature) is drawn from a cold water tank (capacity of 600 liters) using a 0.5 Horsepower pump. The flow was controlled with the help of a bypass value to set the required flow rate in a small range rotameter. For a set of readings, mass flow rates for cold & hot water were kept equal. Both flow rates (cold, hot water) start from 400 LPH and end at 1500 LPH with an increase of 100 LPH for every reading. Two pressure gauges (bourdon tube) were used of a range of 0-2 kg/cm^2 and a least count of 0.001kg/cm^2. One was installed at the entry of the test section and the other just at the exit of the test section. Four PT100 RTDs (Resistance temperature detector) sensors were used for measuring the inlet &outlet temperatures of hot (T_3, T_4) & cold water (T_1, T_2). The detailed overview of the various aspects of the experimental setup is listed in Table 1.



(a)



(b)

Figure 2. (a) Experimental setup, (b) Schematic of double pipe heat exchanger experimental setup.

Table 1. A detailed description of different components of the experimental setup.

Name	Specification	Dimension
Outer G.I. pipe	Outer Diameter	50mm
	Inner Diameter	45mm
	Length	4232mm
Inner copper pipe	Outer Diameter	24.5mm
	Inner Diameter	21.5mm
	Length	4432mm
Flat strip spring insert	Width	2.5mm
	Thickness	1mm
	Inner Diameter	21.5mm
	Length	150mm

2.2 Accretion Techniques Utilized in Current Work

An overview investigation of different types of turbulence generation devices has been conducted in a wide range of Reynolds numbers. A spring turbulator perform better for turbulent flow with conical shape and different arrangements of converging, diverging and converging and diverging. In turbulent flow, conical converging spring shows lower friction factor and disturbance to boundary layer as compared to diverging spring in circular cross-section pipe within range of 10,000 to 34,000 [10]. In contrast, the performance of helical spring tubulators inside the inner pipe of a double pipe heat exchanger. Nusselt number increases as the pitch of helical spring increases at higher Reynold number [27]. The perforated solid metal ring was tested with different open area ratio (0 to 0.0833) in the range of Reynolds numbers 6000 to 12,000 Additionally, it was noted that the friction factor decreased as the perforations increased in the metal ring inserts when liquids used as working fluid [15]. The combined performance of metal rings and twisted tape turbulaors with different pitch ratios (1 and 2) and twist ratios (2,3 and 4) was investigated and found higher heat transfer enhancement but at a cost of higher friction factors in the ranges of Reynold number from 6000 to 24000 [24]. A triple twisted tape utilised as swirl flow generator inserts with four type of twist ratio (1.92 to 6.97) under the condition of constant heat flux. As the twist ratio decreases, values of different parameter (Nu, friction factor and efficiency) increases within the range Reynold number of 7200 to 50,200 [32]. The performance of helical spring tubulators was not as good as that of conical spring tabulators. The perforated metal ring turbulators have close to half the Nusselt number values of conical spring turbulators. Single twisted tape turbulators are found to perform better in flow with a low Reynolds number as opposed to one with a higher Reynolds number because they block the flow, which results in an increased friction factor. Compared to the other swirl inserts, Nusselt number performance of metal rings and twisted tape turbulaors was found to be the best but with higher friction factor value. Hence thermohydraulic performance of twisted tape not better at higher values of Reynold number of turbulent flow. Therefore, it may be estimated that, for double pipe heat exchanger, flat plate turbulator perform well in the turbulent flow because its having all essential benefits of conical spring, metal rings and twisted tape.

In current work, a double pipe heat exchanger is used as a standard configuration. The passive heat transfer enhancement technique was predictable in order to increase the efficiency of the current heat exchanger without affecting the surface area necessary for heat transfer. Flat strip spring turbulators as shown Fig. 3 were used as swirl generators. Diameter of FST was 21.4mm which was only adequately large enough to get fit inside the inner copper tube so that, once inside and after initiation of hot water flow any undue movement or shivering could be prohibited. FST was fabricated in Mohits springs Pvt. Ltd. located in Meerut Utter Pradesh who are specialized in the manufacturing of springs. The FSTs were visualized mounted simultaneously on thin high carbon steel at certain specific gaps known as Pitch. Rods of varying pitches were prepared to insert them in inner copper tubes. The idea behind this concept was to create an obstruction to hot water flow which consequently enhances the turbulence and swirl flow thereby augmenting the rate of heat transfer.

FSTs were mounted by brazing on high carbon steel rod to prevent any flickering when inserted in inner copper tubes and with hot water flow commenced. Several pitches were decided in advance depending upon which the brazing was done. The U bend section was detached by opening the flange couplings, and hence, the FST of P = 15cm was inserted followed by 10cm and 5cm pitches.

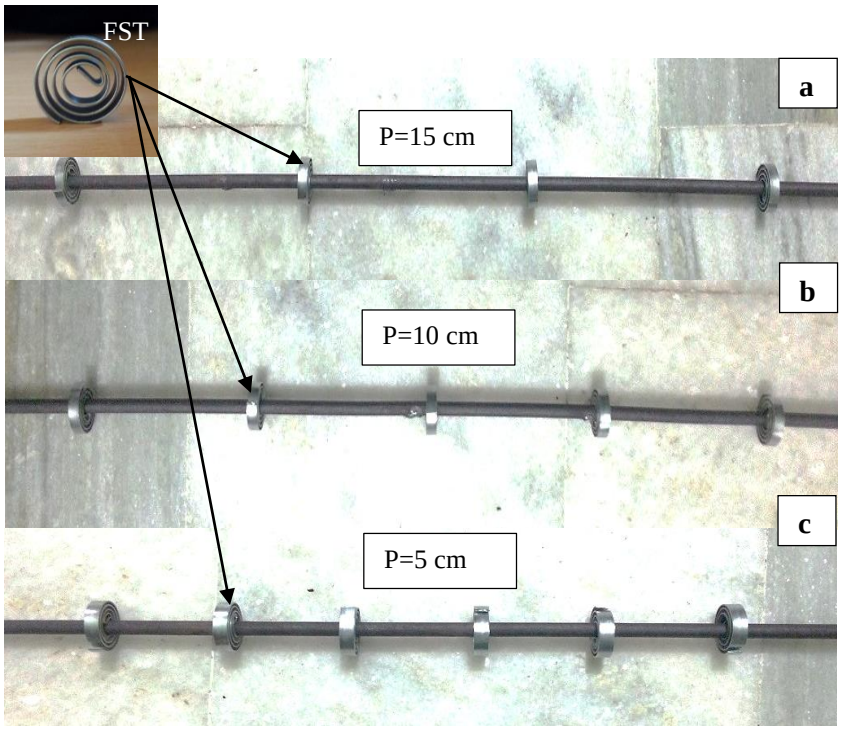


Figure 3. Flat spring tabulator (FST) mounted on rod having (a) 15 cm pitch, (b) 10cm Pitch, (c) 5cm Pitch.

3. Methodology

3.1 Problem Formulation

During the experiments, some numerical investigations constraints/assumptions were taken;
1) Flow is assumed to be steady, non-uniform and incompressible. 2) Neglecting the heat losses in all directions by maintaining isothermal condition 3) the inner side coefficient of thermal expansion and outer side coefficient of thermal contraction of the inner pipe in the concentric pipe of heat exchanger negates each other. 4). the inner surface of the pipe is assumed to be smooth.

3.2 Standardization of the Heat Exchanger Set-up

Calibration of RTD and rotameter is an important task for getting the idea of the accuracy of the experimental setup. For the calibration of RTDs, all of them were dipped (at the same depth) in a water tank maintained at a constant temperature. By taking the reference of one of the RTD (T_4) temperature readings, other RTDs reading values (T_1 to T_3) were corrected accordingly and after taking 8 number of observations, calibration found to be ±1. In the case of rotameter calibration, initially, two buckets (25 kg each) were used for collecting and measuring the weight of water. A flow rate of the small rotameters was varied from 300 LPH to 700 LPH. Each observation was taken for 180 seconds and ound percentage in error ranging from 1.307 to 1.774. And three observations were done for each mass flow rate of cold water rotameters. Similarly, large rotameters calibration was done by varying its mass flow rate from 800-1200 LPH for the same amount of time and observed percentage in error ranging from 1.715 to 2.614.

Before commencing the experimentation earthing of the apparatus was done. Friction factor and pressure drop readings were attained for the plane tube to verify the results with the existing standard equation for friction factor. This procedure was to eliminate the chances of deviation that could occur by repeated insertion and extraction of FSTs, which induce certain scratches (wall roughness) on the inner side of the copper tube. The readings hence obtained were compared with the Blasius equation to verify the plane tube results. This whole procedure was done at a normal temperature.

3.3 Thermal Performance Result and Repeatability

The complete experiment was re-conducted for thermal performance result and repeatability check (shown in Tables 2,3 & 4). For calculating the equivalent Reynolds number, water and pumping power was kept constant.

.

Table 2. Experimental data for repeatability Heat transfer versus Re for FST having pitch 15 cm.

$\dot{m}_{cw}$, $\dot{m}_{hw}$ (Kg/s)	T1 (ºC)	T2 (ºC)	T3 (ºC)	T4 (ºC)	Re_{hw}	Trail 1 Nu_{Exp}	Trail 2 Nu_{Exp}	%diff
0.113	35.9	57	75.8	57.1	9712.452	63.044	59.067	-6.732
0.141	35.8	56.2	75.8	57.9	12043.41	74.622	73.314	-1.783
0.17	35.8	55.8	75.9	58.6	14457.88	87.217	88.222	1.138
0.198	35.9	55.7	75.9	59.6	17115.7	98.551	97.270	-1.317

Table 3. Experimental data for repeatability heat transfer versus Re for FST having pitch 10 cm.

$\dot{m}_{cw}$, $\dot{m}_{hw}$ (Kg/s)	T1 (ºC)	T2 (ºC)	T3 (ºC)	T4 (ºC)	Re_{hw}	Trail 1 Nu_{Exp}	Trail 2 Nu_{Exp}	%diff
0.113	36.2	58.1	76.3	57.1	9643.40	64.32	62.770	-2.482
0.141	36.2	57.7	76.2	57.6	12043.41	78.90	77.715	-1.534
0.17	36.3	56.9	76.2	58.6	14457.88	91.31	92.489	1.269
0.198	36.3	56.7	76.1	59.6	17115.7	107.66	103.88	-3.642

Table 4. Experimental data for repeatability heat transfer versus Re for FST having pitch 5 cm.

$\dot{m}_{cw}$, $\dot{m}_{hw}$ (Kg/s)	T1 (ºC)	T2 (ºC)	T3 (ºC)	T4 (ºC)	Re_{hw}	Trail 1 Nu_{Exp}	Trail 2 Nu_{Exp}	%diff
0.113	36.3	58.7	76.9	57.1	9643.4	68.56	67.28	-1.902497
0.141	36.2	58.3	76.8	57.7	12043.41	83.18	82.11	-1.30313
0.17	36.2	57.6	76.7	58.5	14457.88	95.43	96.76	1.374535
0.198	36.4	57.2	76.5	59.3	17115.7	110.79	106.5	-4.028169

3.4 Factors Affected by Varying Pitch Between Consecutive FST

Pitch over here refers to the distance between two consecutive FST mounted on a brass rod. Apart from the enhancement in heat transfer and thermal performance installation of FST leads in the increase in pumping power requirement. The minimum pitch of the turbulators allows more turbulators on the given length which leads to more friction to the flow, more back pressure generates and all these leads to more pumping power requirement. This arrangement also separates the boundary layer commencing earlier as the pitch between two consecutive turbulators going to be decrease. Fig. 4 represents the effect of FST over variation in pitch ratio

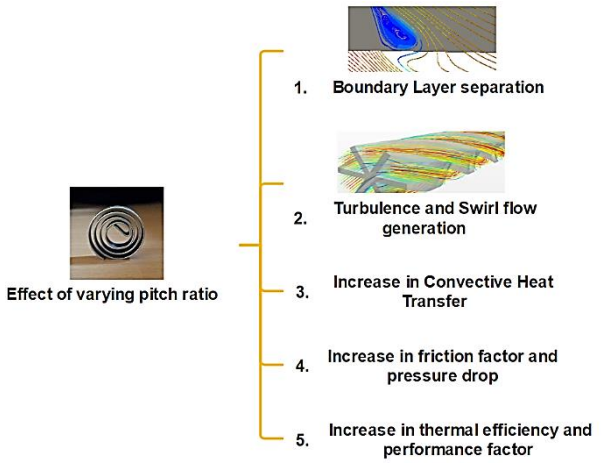


Figure 4. Repercussion by varying pitch value of FST.

4. Data Reduction

Water is taken as a working fluid for all the experiments conducted with a parametric study of the effects of variation in mass flow rate for turbulent case. Installation of FST with varying pitch ratio influences the flow conditions including an adverse effect i.e. increment of friction factor. Different equations required for the basis of such experiments are summed as follows:

Quantity of heat flow [33] for hot and cold water loops can be calculated from Eqn. (1) and (2)

$$\dot{Q}_c = \dot{m}_c \times C_{pc} \times (T_{ce} - T_{ci}) \tag{1}$$

$$\dot{Q}_h = \dot{m}_h \times c_{ph} \times (\mathrm{T}_{hi} - T_{he}) \tag{2}$$

The arithmetic average of heat exchange from the hotter and colder fluid streams is

$$Q_{ave.} = \frac{Q_h + Q_c}{2} \tag{3}$$

The overall heat transfer coefficient can be calculated as

$$U = \frac{Q_{ave.}}{A \times lmtd} \tag{4}$$

$$lmtd = \frac{(T_{hi} - T_{ce}) - (T_{he} - T_{ci})}{\ln\left((T_{hi} - T_{ce}) / (T_{he} - T_{ci})\right)} \tag{5}$$

Where A is the circumferential area of the inner pipe.

4.1 Standard Equation

The non-dimensional numbers; Nu, Re and Pr calculated from Eqn:

$$Nu = \frac{hD}{k} \tag{6}$$

$$\mathrm{Re} = \frac{\rho V D}{\mu} \tag{7}$$

$$\mathrm{Pr} = \frac{\mu c_p}{k} \tag{8}$$

$$\lambda = \frac{\Delta P}{\left(\frac{l}{d}\right)\left(\frac{\rho V^2}{2}\right)} \tag{9}$$

$$V = \frac{\dot{m}}{\rho A_c} \tag{10}$$

4.2 Heat Transfer and Friction Factor Calculation

For the heat transfer calculations, some standard equations from the literatures were used for the experimental flow conditions.

DittusBoelter Equation [34]

$$Nu = 0.023 \mathrm{Re}^{0.8}\, \mathrm{Pr}^{0.3} \tag{11}$$

Friction factor for the different flow arrangements were calculated from Blasius Equation and Darcy-weisbach equations [33]

$$\lambda = \frac{0.3164}{\mathrm{Re}^{0.25}} \tag{12}$$

$$\lambda = \frac{\Delta P}{\frac{L}{D}\frac{\rho V^2}{2}} \tag{13}$$

Thermal performance factor calculation at constant pumping power

$$\left(\lambda \mathrm{Re}^3\right)_{PT} = \left(\lambda \mathrm{Re}^3\right)_T \tag{14}$$

The thermal performance factor and performance evaluation criteria is the key parameter in designing effective heat exchanging devices [35]. The thermal performance factor (η) is the ratio of the Nusselt number ratio (Nu_T/Nu_{PT}) to the friction factor ratio (λ_T/λ_{PT}) considering constant pumping powerand can be represented as

$$\eta = \frac{Nu_T / Nu_{PT}}{\left(\lambda_T / \lambda_{PT}\right)^{1/3}} \tag{15}$$

4.3 Uncertainty Calculation

Mass flow rate, pressure (at inlet and exit), and temperature distribution were the variables measured by the test rig. Before being used in the experimental setting, all variable measuring equipment were calibrated. The two crucial factors, Nusselt number and Reynolds number, were going to be used to understand the experiment's measured variables. Temperature and pressure drop are the most effect-causing variables for Nusselt number, while flow rate is the greatest effect-causing variable for Reynold number. Fluid thermo-physical properties, on the other hand, had a significant impact on both of them.

The root sum square equation, which combines the effects of each distinct input as proposed by Kline and McClintock [36], could be used to determine the uncertainties in the computed findings with improved precision. The uncertainties in the aforementioned equations are caused by inaccuracies in the primary parameters, as indicated in Table 5. The highest computed uncertainty in heat transfer coefficient, Prandtl number, and friction factor obtained from all experiments are 4%, 2.5%, and 5%, respectively. The computed uncertainty in the results is calculated from Eq. 16.

$$R = \pm\sqrt{\left(\frac{\delta R}{\delta x_1}\delta x_1\right)^2 + \left(\frac{\delta R}{\delta x_2}\delta x_2\right)^2 + ...+ \left(\frac{\delta R}{\delta x_n}\delta x_n\right)^2} \quad (16)$$

Where δx is the uncertainty of the independent individual variables and $\frac{\partial R}{\partial x}$ is sensitivity.

Table 5. Uncertainties in the main parameters.

Parameter	Uncertainty (%)
Annulus-side Reynolds number	± 1.77**
Tube-side Reynolds number	± 2.61**
Thermocouple	± 0.1 °C*
Pressure transducer	± 0.35*
Heating wire	± 0.15*

* Based on manufacturer claim.
** Based on calibration.

4.4 Preparation of Wilson Chart and Standard Equations

The calculation of the film heat transfer [37], [38] can be estimated by the very popular technique known as the Wilson plot. This plot is based on the overall thermal resistance in the total heat transfer in the form of convection. For all sets of experiment, except the first term, all other resistance is constant on the RHS of Eq. 17.

$$\frac{1}{U_i} = \frac{1}{h_i} + \frac{h_i}{d_o h_o} + \frac{x_w d_i}{k_w d_i} \quad (17)$$

As a flow on inner waterside is turbulent and the variation in thermal properties are negligible, for a smooth tube with a flow rate of more than 10000 Reynold number, the Seider Tate equation is of the form

$$h_i = A\mathrm{Re}^{0.8} \quad (18)$$

Therefore, from Eq. 17 and Eq. 18, it can be written as

$$\frac{1}{U_i} = \frac{1}{A\mathrm{Re}^{0.8}} + K \quad (19)$$

Here, K is a constant and its value is to be estimated on the intercept of the y-axis of the Wilson chart (1/Ui vs. $1/\mathrm{Re}^{0.8}$). This value was put in Eq. 17 to obtain the value of h_i. This expression is in the form of the exponent of Reynold number.

The same procedure as was done for the plane tube was repeated for whole pitches of FST. Heat transfer results were obtained by preparing Wilson Charts between Reynolds number (Re) and overall heat transfer coefficient (U) it can be seen in section 5.3.

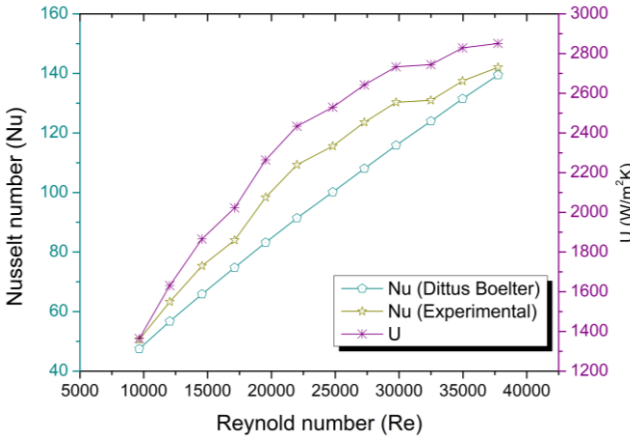
5. Result and Discussion

In this analysis effect of Flat Strip Spring Turbulator (FST) installed inside the double pipe heat exchanger in the counter-current flow arrangement was analysed. Friction factor increment means a decrease in the pitch ratio of turbulators used in the experiment, which ultimately affects the pressure drop i.e. pumping power. But, for maintaining higher value of Reynolds number, a constant (high) pumping power is required. Hence, it is desired to maintain an optimum condition i.e. balance between pitch ratio, friction factor, pressure loss, and ultimately in pumping power. The main function of a turbulator is to generate detachment and reattachment of flow around them and as a result of this, reduces the effect of the boundary layer (when heat is transferred through the laminar effect). This separation of flow generates proper mixing regions (with large turbulence energy) which ultimately destroy the eddies and vortex formation for heat transfer enhancement.

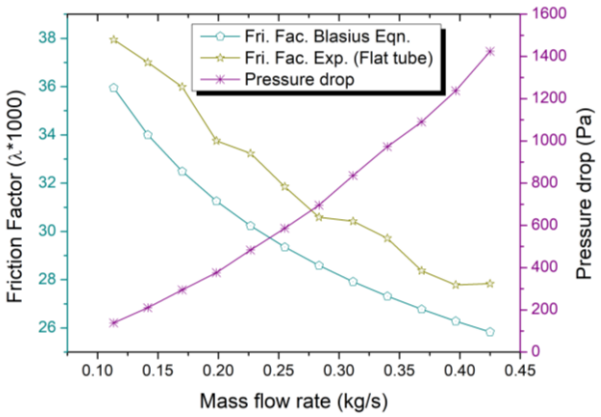
5.1 Plain Tube Experimentation

For hot water, heaters were used to heat the water to 76°C, as this temperature lied within the limits of the heating capacity of heaters and could be easily maintained in the working conditions. A temperature regulator is used for maintaining constant temperature. The tank is connected to pump (centrifugal type) for circulation of hot water inside the inner pipe of the heat exchanger and for controlling the flow rate a bypass valve is used (recirculation the hot water back into the tank). This hot water allowed to pass through the inner tube of heat exchanger at a desired mass flow rate and simultaneously cold water flow was initiated and this set up was left continued for at least 20 minutes to attain a steady-state condition. Now that a steady state was achieved, the main experimentation was commenced by adjusting the mass flow rates by using rotameters of both cold water and hot water side at desired values. The readings from RTDs were obtained only when a steady condition is achieved. This procedure was repeated at various flow rates (0.1134-0.425 Kg/s) of cold and hot water. Under the assumption of uniform heat flux, the calculation of Nusselt numbers is done and compared with the fundamental Eqn. 11 (provided by Dittus and Boelter) for validating the current plane tube.

The Nusselt number variation with Reynolds number for plain tube experiment are shown in the Fig.5(a).Whereas, it was noted that the variation of Nusselt number within the acceptable range of Dittus– Boelter equation correlation [34]. Before beginning any heat exchanger related experiment utilizing turbulator, measurement of friction factor as a secondary parameter is necessary. So, verifying (under similar test conditions) the current plain inner copper tube for the friction factor parameter by comparing of the current data obtained with those obtained from the Blasius correlation. This variation of friction factor with the Blasius equation is show in Fig.5(b).



(a)



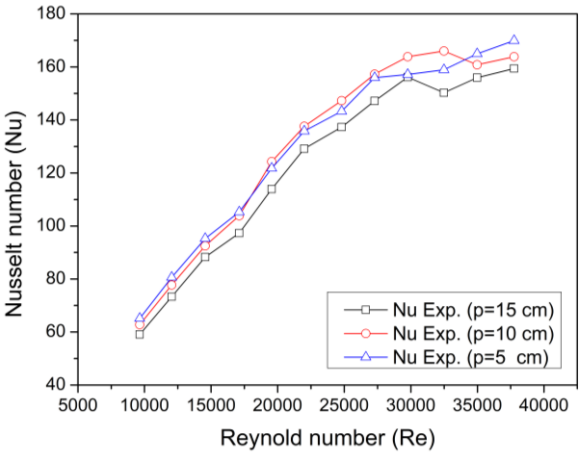
(b)

Figure 5. (a) Nusselt number and overall heat transfer coefficient for plane tube, (b) Variation in friction factor and pressure drop for plane tube.

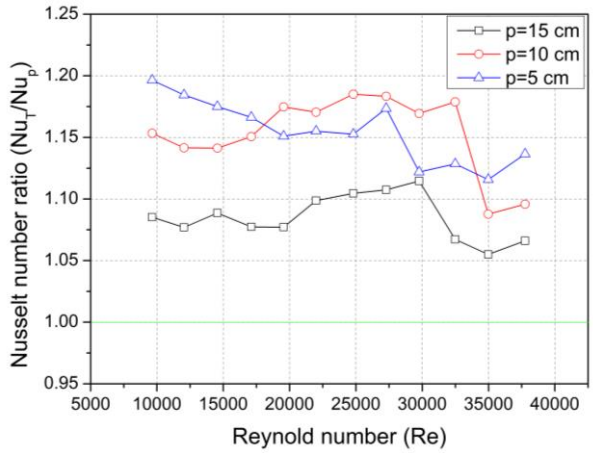
5.2 Effect of FST Turbulator on Nusselt Number

Effect on tubes with FST at different pitches (PR=15, 10, and 5cm) on the heat transfer rate (Nu) is demonstrated in Fig. 6. For all values of Reynolds number, the heat transfer rate for FST is higher than those of the plain tube and this is due to the destruction of the thermal boundary layer present near the inner tube wall. In different turbulator arrangements, it is found that with the decrease of pitch the heat transfer rate increases. This is because since, FST with a smaller PR of 5cm, create more intervention in the generation of thermal as well as hydrodynamic boundary layer with a greater degree of turbulence than that of higher PR of 15cm. The quantitative analysis reveals that the percentage heat transfer rate in the tube with FST of P =15cm, P=10cm, P= 5cm is 7.93%, 13.09%, and 14.26% higher than those in the plain tube at Re =17115.70.

Fig. 6. (b) Indicate the Nusselt number ratio variation with respect to Reynolds number. From the graph, it is evident that Nusselt number values decrease with increasing Reynolds number or its having higher rate of heat transfer at lower values of Reynolds number. This occurs primarily because at low Reynolds numbers the thermal boundary layer thickness is higher near the surface of the pipe, limiting heat transfer, while after putting the FST in place, the boundary layer effect is no longer noticeable. It is evident that, 10 cm pitch FST has a higher value of Nusselt number ratio as compared to 15cm pitch. In all three cases, the values attained were greater than unity, demonstrating the advantage of using FST as an insert in heat exchangers as compared to a plan tube. It is found that the highest and lowest values of the Nusselt number ratio are in between 1.20 and 1.13 for FSTs with a pitch of 5 cm, whereas in case of pitch 15 cm it is between 1.10 and 1.05 for FSTs.



(a)



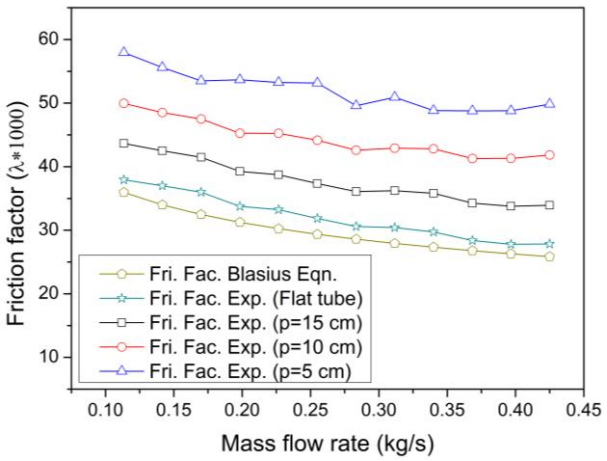
(b)

Figure 6. (a) Heat transfer enhancement by FST, (b) Effect of pitch values on Nusselt number.

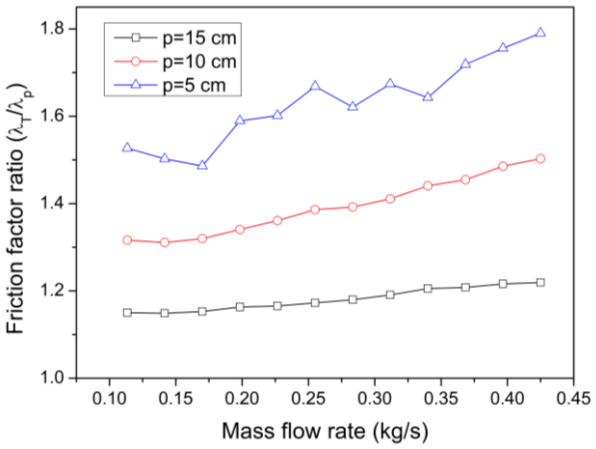
5.3 Effect of FST Turbulator on Friction Factor

The friction factor is one of the important parameters for improving the thermal performance factor in heat exchangers. Across the test section, friction factor value varies directly with the pressure drop values and inversely with the square of fluid velocity values. Whereas, by considering Reynolds number increment, friction factor decreases but pressure drop increases. So, as the turbulator used in the heat exchanger, the friction factor will increase because of the obstruction in the flow. Therefore, it is very important to find out a turbulator that imparted the lowest friction factor.

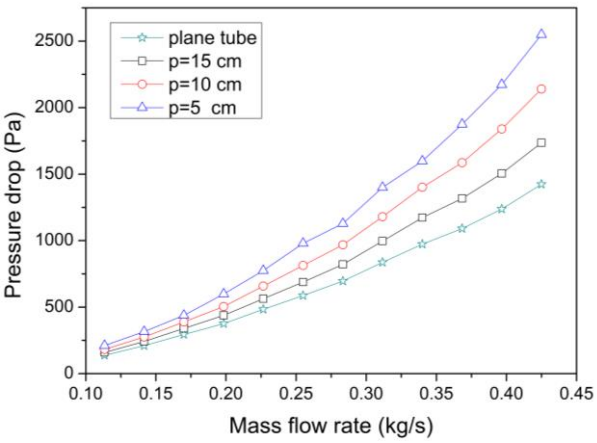
From the scenario of Fig. 7 (a&b), It is concluded that the value of friction factor increases as the pitch ratio decreases for the same Reynold number. The reason behind this trend is smaller the distance between FSTs, and more obstructions faced by the flow which ultimately causes an increase in friction factor. The simple fact is that, when the distance between two consecutive FSTs is less, more space is available for mounting the FSTs on the cylindrical rod inside the inner copper tube, thus the more obstruction against the flow stream of hot water. Hence, more turbulence results in a high-pressure drop.



(a)



(b)



(c)

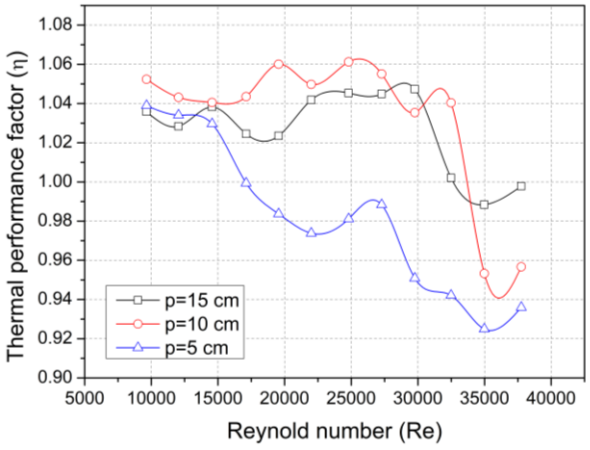
Figure 7. (a) Influence of different pitch value on (a-b) turbulator on friction factor, (c) Pressure drop.

The turbulator acts as an obstruction, as the difference between two consecutive FST increases, local flow velocity quantity also increases (which means local Reynolds number increase). The occurrence of these events may generate lots of vortexes and these vortexes when faces centrifugal force due to secondary flows cause pressure drop as compare to plan tube. As shown in Fig. 7(c) that as the FST pitch decreases pressure drop increases for the constant value of the Reynolds number. FST pitch increment means that obstruction located at larger distances and as a result of this decrement in Reynolds number and generation of vortexes is observed with an increment in hydraulic diameter. All these phenomena together cause the decrement in pressure drop.

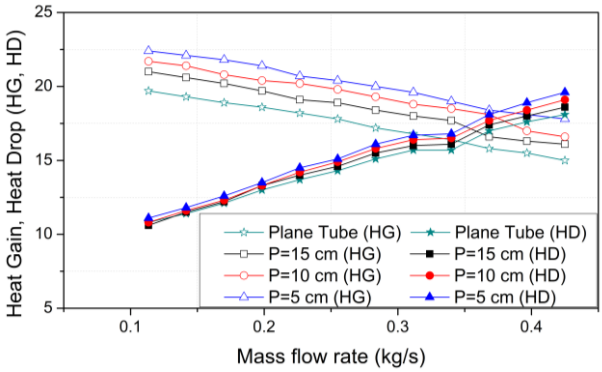
5.4 Effect of FST Turbulator on the Thermal Performance Factor

Only thermal aspects cannot be set as the final criteria for the selection of FST. Here, an optimal value of pitch should be selected which has better heat transport capability and also have a lower value of frictional loss (pumping power). The thermal performance factor is also one of the important factors to count both the effect in the heat exchanger. A significant increment in Nusselt number and friction factor is always observed for various types of turbulators. So, designing the geometry of turbulator is an important task with a view of obtaining maximum thermal performance factors. As can be seen in Fig. 8(a) at constant pumping power, it is evident from the graph that Reynold number and thermal performance factor are in inverse relation with each other. Also, this can be concluded that because of improved thermal and hydraulic performance, the FSTs at P = 10 cm (having the same pumping power) are most efficient.

As can be seen from Fig. 8(b) there occur a significant change in heat gain and heat drop trends at different mass flow rate. It was observed that at higher mass flow rate heat drop and heat gain become approximately equal and the general trend was the heat drop rate decreases with increase in mass flow rate and rate of heat gain increases with increase in mass flow rate.



(a)



(b)

Figure 8. (a) Variation in thermal performance factor with Reynold number, (b) HT characteristics at constant $\dot{m}_{hw}$ & $\dot{m}_{cw}$.

5.5 Experimental Results Conjugated to Wilson Plot

Wilson plot is a very important tool that is utilized to check the performance of different types of heat exchangers [42]. Through this approach, the overall heat transfer coefficient value can directly be calculated which is not

easier indirect calculation from the experimental results due to inaccessible surface temperature values. The experimental results obtained from the experiment on double pipe heat exchanger experimental setup with flat strip spring Turbulator insert were analyzed with Wilson chart (having variation between overall heat transfer coefficient (U) and Reynold number). The results were plotted between 10000/Ui and $10000/Re^{0.8}$, which is presented in Fig. 9. The variation of 10000/U follows the linear trend as of the Wilson plot. However, the experimental values are slightly lesser than that of the straight-line of the Wilson plot.

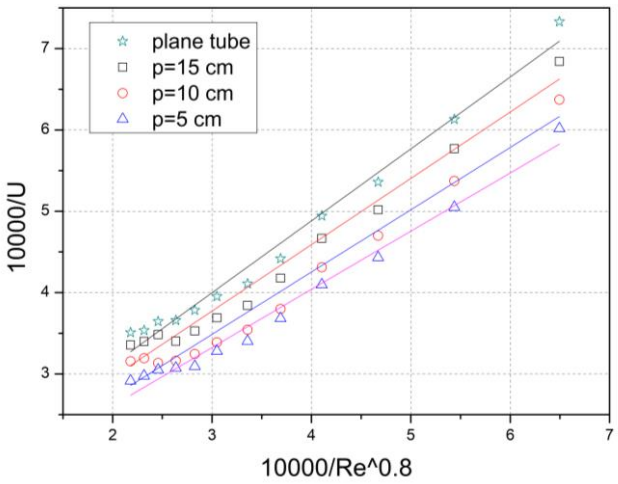


Figure 9. Visualization of overall heat transfer coefficient with Wilson plot.

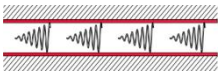
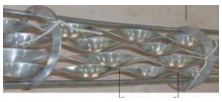
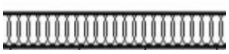
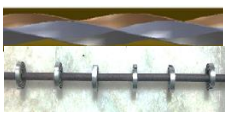
5.6 Comparison of Performance Parameters of Different Turbulators

In the Table 6, the term "Nusselt number ratio" (Nu_{max}/Nup) is the ratio of maximum Nusselt number after augmentation to Nuseelt number of the plane tube, and "friction factor ratio" ($\lambda_{max}/\lambda p$) is the ratio of maximum friction factor after augmentation to friction factor of the plan tube. In current work, the Nusselt number ratio comes only 1.2 at the expense of friction factor ratio of 1.44 in a range of Reynolds no. (9000–38,000). Whereas, maximum increment in Nusselt number ratio is claimed by Singh et. al. [24] i.e. 4.6 but at a higher value of friction factor ratio of 36.07 in a range of Reynolds no. (6300– 22500) and as it is well known that, with the increment in Reynolds number friction factor also increases which further increases required pumping power and makes the heat exchange process uneconomical. So, in the present study even at a higher value of Reynolds number friction factor obtaining well-accepted range as compared to the previous studies.

6. Conclusion

The present experimental study presents the potential application of FST to enhance the heat transport performance of a concentric double tube heat exchanger. The experimental objectives investigation was successfully carried out with the insertion of FST at various pitch values at varying cold and hot water flow rate of 500 to 1500 LPH and 500 to 2000 LPH, respectively. All the time, the experiment was in turbulent flow with Reynold number ranging from 9000 to 38000, which significantly influences the different parameter of heat exchanger (Nusselt number, friction, and thermal performance factor). The most remarkable conclusions drawn after conducting this experiment are:

Table 6. Comparison of performance parameters of different turbulators.

Author	Turbulator used	Parameters	Reynolds Number	Nusselt Number ratio	Friction factor ratio	Image
Karakaya et. al. [10]	Conical spring turbulators	different cone angle (30°, 45°, 60°)	10,000 to 34,000	$Nu_{max} \approx 3.33\ Nu_p$	$\lambda_{max} \approx 1.72\ \lambda_p$	
Sheikholeslami et. al. [15]	typical and perforated circular turbulators	open area ratio (0 to 0.0625), pitch ratio (1.83 to 5.83)	6000 to12,000	$Nu_{max} \approx 2.12\ Nu_p$	$\lambda_{max} \approx 11.40\ \lambda_p$	
Singh et. al. [24]	solid ring tubular (SRT) with Number of twisted tapes (TT): 1, 2, 3, 4	Pitch ratios (1 and 2_, twist ratios (2, 3, and 4)	6300 to 22500	$Nu_{max} \approx 4.6\ Nu_p$	$\lambda_{max} \approx 36.07\ \lambda_p$	
Akpinar[27]	helical (spring shaped)		6500 to 13,000	$Nu_{max} \approx 2.64\ Nu_p$	$\lambda_{max} \approx 2.74\ \lambda_p$	
Bhuiya et al. [39]	Triple twisted tape	Twist ratio-1.92–6.79	7200 to 50,200	$Nu_{max} \approx 3.85\ Nu_p$	$\lambda_{max} \approx 4.2\ \lambda_p$	
Present study	Flat strip spring turbulator	At different pitch ratio of 5, 10 &15	9000 to 38000	$Nu_{max} \approx 1.2\ Nu_p$	$\lambda_{max} \approx 1.44\ \lambda_p$	

- The enhancement of heat transfers with flat spring inserts of P =15cm, P=10cm, P= 5cm is 7.93%, 13.09%, and 14.26% higher than those in the plain tube at Re =17115.69 and flow rates at 0.19839 kg/s. Over the range investigated, the maximum heat transfer rise was about 20% for FST at pitch 5cm.
- The heat transfer rate and friction factor of FSTs increase with decreasing pitch due to the number FSTs present on the cylindrical rod was high inserted in the copper tube. However, the thermal performance factor increases with decreasing pitch
- The friction factor obtained from using the PCR with pitch 15, 10, and 5cm are found to be respectively, 17%, 30%, and 39%, over the plain tube at Reynolds number of 17997.71. An increase in the pitch of FSTs causes a reduction in Nusselt numbers as well as friction factors. The maximum friction factor rise was about 44% for Reynolds number of 22497.1.
- The maximum heat gain & heat drop-in plan tube is 19.7 & 18.1. On 15cm pitch, Maximum heat gain& heat drop is 21 & 18.6. On 10cm pitch, Maximum heat gain & heat drop was 21.7 & 19.1. On 5cm pitch, Maximum heat gain & heat drop is 22.4 & 19.6.

As the pitch between two consecutive FSTs decreases not only heat transfer rate increase but also an adverse effect increases i.e. friction factor. An optimal pitch value is suggested from this study because increasing the turbulator is not only the solution for heat transfer enhancement, a downside of this arrangement is the increasing pressure drop which leads to more power requirement and this should be considered for optimization to the maximum performance of heat exchanger.

Nomenclature

L	Total length of pipe
D	Hydraulic diameter
V	Velocity of water
ΔP	Frictional pressure drop
c	Specific heat
d	Diameter of inner pipe
Nu	Nusselt Number
Re	Reynold Number
P	Pitch
ṁ	Mass flow rate
c	Specific heat
T	Turbultor
A	Area
PT	Plane Tube
λ	Friction factor
ρ	Density of fluid
η	Thermal performance factor
LMTD	Log Mean Temperature Difference
RTD	Resistance temperature detector
HG	Heat Gain
HD	Heat Drop

Subscript	
i	Inner
o	Outer
c	Cross sectional
max	Maximum

References:

[1] A. E. Bergles, R. L. Bunn, and G. H. Junkhan, "Extended performance evaluation criteria for enhanced heat transfer surfaces," *Letters in Heat and Mass Transfer*, vol. 1, no. 2, pp. 113–120, Nov. 1974, doi: 10.1016/0094-4548(74)90147-7.

[2] W. M. Rohsenow, J. P. Hartnett, E. N. Ganic, and P. D. Richardson, *Handbook of Heat Transfer Fundamentals (Second Edition)*, vol. 53, no. 1. 1986.

[3] D. B. Berkowitz, *Handbook on Syntheses of Amino Acids: General Routes for the Syntheses of Amino Acids*, vol. 132, no. 50. 2010.

[4] H. S. Dizaji and S. Jafarmadar, "Experiments on New Arrangements of Convex and Concave Corrugated Tubes through a Double-pipe Heat Exchanger," *Experimental Heat Transfer*, vol. 29, no. 5, pp. 577–592, Sep. 2016, doi: 10.1080/08916152.2015.1046015.

[5] H. M. Şahin, E. Baysal, and A. R. Dal, "Experimental and numerical investigation of thermal characteristics of a novel concentric type tube heat exchanger with turbulators," *International Journal of Energy Research*, vol. 37, no. 9, pp. 1088–1102, Jul. 2013, doi: 10.1002/er.2919.

[6] J. K. Dasmahapatra and M. R. Rao, "Laminar flow heat transfer to generalised power law fluids inside circular tubes fitted with regularly spaced twisted tape elements for uniform wall temperature condition," in *American Society of Mechanical Engineers, Heat Transfer Division, (Publication) HTD*, 1991, vol. 174, pp. 51–58.

[7] S. Al-Fahed and W. Chakroun, "Effect of tube-tape clearance on heat transfer for fully developed turbulent flow in a horizontal isothermal tube," *International Journal of Heat and Fluid Flow*, vol. 17, no. 2, pp. 173–178, Apr. 1996, doi: 10.1016/0142-727X(95)00096-9.

[8] R. M. Manglik and A. E. Bergles, "Heat transfer enhancement and pressure drop in viscous liquid flows in isothermal tubes with twisted-tape inserts," *Wärme- und Stoffübertragung*, vol. 27, no. 4, pp. 249–257, Apr. 1992, doi: 10.1007/BF01589923.

[9] P. Zamankhan, "Heat transfer in counterflow heat exchangers with helical turbulators," *Communications in Nonlinear Science and Numerical Simulation*, vol. 15, no. 10, pp. 2894–2907, Oct. 2010, doi: 10.1016/j.cnsns.2009.10.025.

[10] H. Karakaya and A. Durmuş, "Heat transfer and exergy loss in conical spring turbulators," *International Journal of Heat and Mass Transfer*, vol. 60, no. 1, pp. 756–762, May 2013, doi: 10.1016/j.ijheatmasstransfer.2013.01.054.

[11]D. Yadav, Z. Upadhyay, A. Kushwaha, and A. Mishra, "Analysis Over Trio-Tube with Dual Thermal Communication Surface Heat Exchanger [T.T.H.Xr.]," in *Lecture Notes in Mechanical Engineering*, 2020, pp. 1–13.

[12]D. Yadav, A. Kushwaha, and A. Mishra, "Design and Fabrication of Trio Tube Heat Exchanger Experimental Setup," *SSRN Electronic Journal*, 2020, doi: 10.2139/ssrn.3576468.

[13]M. Sheikholeslami and D. D. D. Ganji, "Heat transfer enhancement in an air to water heat exchanger with discontinuous helical turbulators; experimental and numerical studies," *Energy*, vol. 116, pp. 341–352, Dec. 2016, doi: 10.1016/j.energy.2016.09.120.

[14]M. Sheikholeslami, M. Gorji-Bandpy, and D. D. D. Ganji, "Effect of discontinuous helical turbulators on heat transfer characteristics of double pipe water to air heat exchanger," *Energy Conversion and Management*, vol. 118, pp. 75–87, Jun. 2016, doi: 10.1016/j.enconman.2016.03.080.

[15]M. Sheikholeslami and D. D. D. Ganji, "Heat transfer improvement in a double pipe heat exchanger by means of perforated turbulators," *Energy Conversion and Management*, vol. 127, pp. 112–123, Nov. 2016, doi: 10.1016/j.enconman.2016.08.090.

[16]K. Nanan, C. Thianpong, M. Pimsarn, V. Chuwattanakul, and S. Eiamsa-ard, "Flow and thermal mechanisms in a heat exchanger tube inserted with twisted cross-baffle turbulators," *Applied Thermal Engineering*, vol. 114, pp. 130–147, Mar. 2017, doi: 10.1016/j.applthermaleng.2016.11.153.

[17]N. Mashoofi, S. M. Pesteei, A. Moosavi, and H. Sadighi Dizaji, "Fabrication method and thermal-frictional behavior of a tube-in-tube helically coiled heat exchanger which contains turbulator," *Applied Thermal Engineering*, vol. 111, pp. 1008–1015, Jan. 2017, doi: 10.1016/j.applthermaleng.2016.09.163.

[18]D. Panahi *et al.*, "Heat transfer enhancement of shell-and-coiled tube heat exchanger utilizing helical wire turbulator," *Applied Thermal Engineering*, vol. 115, no. 2–3, pp. 607–615, Mar. 2017, doi: 10.1016/j.applthermaleng.2016.12.128.

[19]S. P. Nalavade, C. L. Prabhune, and N. K. Sane, "Effect of novel flow divider type turbulators on fluid flow and heat transfer," *Thermal Science and Engineering Progress*, vol. 9, pp. 322–331, Mar. 2019, doi: 10.1016/j.tsep.2018.12.004.

[20]S. Khorasani, S. Jafarmadar, S. Pourhedayat, M. A. A. Abdollahi, and A. Heydarpour, "Experimental investigations on the effect of geometrical properties of helical wire turbulators on thermal performance of a helically coiled tube," *Applied Thermal Engineering*, vol. 147, pp. 983–990, Jan. 2019, doi: 10.1016/j.applthermaleng.2018.09.092.

[21]A. E. Zohir, M. A. Habib, and M. A. Nemitallah, "Heat Transfer Characteristics in a Double-Pipe Heat Exchanger Equipped with Coiled Circular Wires," *Experimental Heat Transfer*, vol. 28, no. 6, pp. 531–545, Nov. 2015, doi: 10.1080/08916152.2014.915271.

[22]N. Budak, H. L. Yucel, and Z. Argunhan, "Experimental and Numerical Investigation of the Effect of Turbulator on Heat Transfer in a Concentric-type Heat Exchanger," *Experimental Heat Transfer*, vol. 29, no. 3, pp. 322–336, May 2016, doi: 10.1080/08916152.2014.976723.

[23]A. Kumar, S. Chamoli, M. Kumar, and S. Singh, "Experimental investigation on thermal performance and fluid flow characteristics in circular cylindrical tube with circular perforated ring inserts," *Experimental Thermal and Fluid Science*, vol. 79, pp. 168–174, Dec. 2016, doi: 10.1016/j.expthermflusci.2016.07.002.

[24]V. Singh, S. Chamoli, M. Kumar, and A. Kumar, "Heat transfer and fluid flow characteristics of heat exchanger tube with multiple twisted tapes and solid rings inserts," *Chemical Engineering and Processing: Process Intensification*, vol. 102, pp. 156–168, Apr. 2016, doi: 10.1016/j.cep.2016.01.013.

[25]A. Kumar, S. Singh, S. Chamoli, and M. Kumar, "Experimental Investigation on Thermo-Hydraulic Performance of Heat Exchanger Tube with Solid and Perforated Circular Disk Along with Twisted Tape Insert," *Heat Transfer Engineering*, vol. 40, no. 8, pp. 616–626, May 2019, doi: 10.1080/01457632.2018.1436618.

[26]R. Datt, M. S. Bhist, A. D. Kothiyal, R. Maithani, and A. Kumar, "Effect of square wing with combined solid ring twisted tape inserts on heat transfer and fluid flow of a circular tube heat exchanger," *International Journal of Green Energy*, vol. 15, no. 12, pp. 663–680, Sep. 2018, doi: 10.1080/15435075.2018.1525552.

[27]E. K. Akpinar, "Evaluation of heat transfer and exergy loss in a concentric double pipe exchanger equipped with helical wires," *Energy Conversion and Management*, vol. 47, no. 18–19, pp. 3473–3486, Nov. 2006, doi: 10.1016/j.enconman.2005.12.014.

[28]C. Maradiya, J. Vadher, and R. Agarwal, "The heat transfer enhancement techniques and their Thermal Performance Factor," *Beni-Suef University Journal of Basic and Applied Sciences*, vol. 7, no. 1, pp. 1–21, Mar. 2018, doi: 10.1016/j.bjbas.2017.10.001.

[29]M. Sheikholeslami, M. Gorji-Bandpy, and D. D. Ganji, "Review of heat transfer enhancement methods: Focus on passive methods using swirl flow devices," *Renewable and Sustainable Energy Reviews*, vol. 49, pp. 444–469, Sep. 2015, doi: 10.1016/j.rser.2015.04.113.

[30]M. Omidi, M. Farhadi, and M. Jafari, "A comprehensive review on double pipe heat exchangers," *Applied Thermal Engineering*, vol. 110, pp. 1075–1090, Jan. 2017, doi: 10.1016/j.applthermaleng.2016.09.027.

[31]H. Li *et al.*, "A comprehensive review of heat transfer enhancement and flow characteristics in the concentric pipe heat exchanger," *Powder Technology*, vol. 397, p. 117037, Jan. 2022, doi: 10.1016/j.powtec.2021.117037.

[32]M. M. K. Bhuiya, M. S. U. Chowdhury, M. Shahabuddin, M. Saha, and L. A. Memon, "Thermal characteristics in a heat exchanger tube fitted with triple twisted tape inserts," *International Communications in Heat and Mass Transfer*, vol. 48, pp. 124–132, Nov. 2013, doi: 10.1016/j.icheatmasstransfer.2013.08.024.

[33]C. Ringsted, K. Eliasen, I. H. Gøthgen, and O. Siggaard-Andersen, *Positive correlation between "the arterial oxygen extraction tension" and mixed venous po2 but lack of correlation between "the oxygen compensation factor" and cardiac output in 38 patients*, vol. 50, no. S203. 1990.

[34]F. W. Dittus and L. M. K. Boelter, "Heat transfer in automobile radiators of the tubular type," *International Communications in Heat and Mass Transfer*, vol. 12, no.

1, pp. 3–22, 1985, doi: 10.1016/0735-1933(85)90003-X.

[35]R. L. Webb, "Performance evaluation criteria for use of enhanced heat transfer surfaces in heat exchanger design," *International Journal of Heat and Mass Transfer*, vol. 24, no. 4, pp. 715–726, Apr. 1981, doi: 10.1016/0017-9310(81)90015-6.

[36]S. Kline and F. McClintock, "Describing uncertainties in single-sample experiments," *Mechanical Engineering*, vol. 75, pp. 3–8, 1953.

[37]D. Wilkie, "Wilson Plot," in *A-to-Z Guide to Thermodynamics, Heat and Mass Transfer, and Fluids Engineering*, Begellhouse, 2011.

[38]J. Fernández-Seara, F. J. Uhía, J. Sieres, and A. Campo, "Experimental apparatus for measuring heat transfer coefficients by the Wilson plot method," *European Journal of Physics*, vol. 26, no. 3, pp. N1–N11, May 2005, doi: 10.1088/0143-0807/26/3/N01.

[39]M. M. Bhunia, K. Panigrahi, S. Das, K. K. Chattopadhyay, and P. Chattopadhyay, "Amorphous graphene – Transformer oil nanofluids with superior thermal and insulating properties," *Carbon*, vol. 139, pp. 1010–1019, Nov. 2018, doi: 10.1016/j.carbon.2018.08.012.

Ethical Hacking of IEEE 802.11 Encryption Protocols

Sandesh Jain, Sarthak Pruthi, Vivek Yadav

Department of Information Technology
Delhi Technological University, Delhi,India

Abstract - **The widespread usage of smart terminals such as smartphones, which provide significant convenience in our daily lives, has resulted in an increase in information security issues. Researchers are increasingly concerned about network security and calculations. As more devices join a wireless network, important and sensitive data is transferred over the air between users, making data packets easy to sniff and grab. The purpose of this project is to perform penetration testing to find flaws in the WEP, WPA, and 802.11i WPA2, WPA3 security protocols, as well as to conduct anonymous attacks against them using the macchanger script. Penetration testing in WEP was carried out using Kali Linux and its Aircrack-ng tools, which exploited a 4-way handshake for WPA/WPA2.**

Index Terms— **Encryption, Protocols, WPA, WPA2, four-way handshake, KRACK, WPA3.**

I. INTRODUCTION

Wireless networks are one of the more recent innovations that the internet has brought into our lives. Wireless technology is the mechanism of sharing information using invisible waves in the air using electromagnetic or acoustic waves. The challenge of information security is becoming increasingly critical as wireless network technology develops and gets more widely used.

Wireless Fidelity (WiFi) is a modern wireless network model defined by the IEEE 802.11 standard. A hostile activity intended targeting wireless system information or wireless networks is known as a wireless attack. The wireless network has several faults of its own. It uses radio waves to send signals and must first establish a connection before being used. A penetration test is a malicious attack on a target system that achieves access control by emulating an attacker's techniques and methods with the client's legal authorization; it is a test method for evaluating information system security control measures. This paper provides a WiFi penetration test method based on Kali Linux that uses methods such as monitoring, sniffing, capturing, data analysis, WiFi password cracking, pseudo-wireless access point spoofing, and other techniques to improve the security of WiFi networks.

II. LITERATURE REVIEW

In most residential and business networks, the IEEE 802.11 standard defines WLAN characteristics; 802 deals with LAN and MAN, while.11 deals with WLAN. The available radio frequency spectrum varies significantly depending on the regulatory domain. Collision-avoidance carrier-sense multiple access and medium access control are used in the 802.11 protocol family, which means that equipment listens for other users on a channel before transmitting each frame. The first edition of the standard was released in 1997. The protocols are typically used in conjunction with IEEE 802.2 to carry Internet Protocol traffic. 802.11a, 802.11b, and 802.11g are the most extensively used and supported standards.

III. TOOLS AND TECHNOLOGIES USED

Tools and Technologies used to implement and perform working model to wifi encryption were:

- Wifi Adapter
- Raspberry pi kit
- Wifi Router
- Raspberry Pi (Power Adapter)
- Bluetooth Adapter
- Kali Linux
- Aircrack-ng, Airodump-ng, Airplay-ng

IV. WLAN PROTOCOLS

A. Working on WLAN Protocols

1. Wired Equivalent Privacy (WEP)

Wired Equivalent Privacy(WEP) is a security protocol which was introduced as part of IEEE 802.11 standard in 1997. It uses an RC4 encryption algorithm to encrypt the plain text and unencrypted integrity check value to ensure the integrity of the plain text when it is transmitted from one end to the other. It operates at the data link layer and the physical layer. In 2001 major flaws such as short IV size, keystream reuse which proved that data transmitted can be easily captured and

tampered. With these flaws WPA was introduced to remove the vulnerability of weak encryption techniques.

2. Wireless Protected Access(WPA)

WPA was created as a WiFi security protocol. It's similar to WEP, but it encrypts data using a temporal key integrity technique (TKIP). To avoid the attacks that WEP allows, TKIP provides a new 128-bit key for each packet. Users can upgrade to TKIP from earlier WLAN equipment without changing hardware because TKIP comprises several methods that encapsulate WEP. Message integrity Check (MIC), IV sequencing mechanism, Per-packet key mixing function, and Re-keying mechanism are four additional algorithms included in TKIP to boost key strength.

A per-packet key mixing mechanism is used to improve cryptographic strength. A re-keying approach is employed to generate a new key for every 10,000 packets. A hashing-based initialization-vector sequencing technique is used. WPA uses TKIP, which dynamically changes the encryption key used by the computers, preventing intruders from matching the secure network's encryption key. A message authentication code (MAC) is a cryptographic way of confirming that communications have not been tampered with. WPA uses the Extensible Authentication Protocol (EAP) to authenticate computers rather than relying exclusively on the plaintext of their MAC address.

3. Wireless Protected Access2(WPA2)

Although Wi-Fi signals are broadcast in the air and can be readily intercepted, encrypting wireless data is critical for security. WPA II is an 802.11 wireless security standard that employs 128-bit encryption and passwords to prevent unauthorized access to critical information.

This protocol uses a single pass-key (PSK) that all devices and the Access Point share for network authentication. The PSK can be 8 to 63 characters long. An attacker can gain access to the network if he discovers this one-of-a-kind PSK. Every device develops and maintains a PMK based on the PSK or the AP name until it changes. When a client attempts to connect to an authenticator, the 4-way handshake procedure begins, and a Pairwise Transient Key (PTK) is generated, which is used to encrypt data between a client and an access point and is changed at least once every 65,535 packets.

Pairwise Master Key Generation

Using the function below, all devices calculate PMK from PSK. The data is encoded using HMAC-SHA1 by the key derivation function PBKDF2. These routines are used to reduce vulnerabilities to brute force attacks because of their high computational cost.

PMK = PBKDF2(HMAC-SHA1, PSK, SSID, 4096, 256)

B. Comparison of Different WLAN Protocols

	WEP	WPA	WPA2	WPA3
RELEASE YEAR	1999	2003	2004	2018
ENCRYPTION METHOD	Rivest Cipher 4(RC4)	Temporal Key Integrity Protocol(TKIP) with RC4	CCMP and Advanced Encryption Standard	Advanced Encryption Standard (AES)
SESSION KEY SIZE	40-bit	128-bit	128-bit	128-bit(Personal) 192-bit(Enterprise)
CIPHER TYPE	Stream	Stream	Block	Block
DATA INTEGRITY	CRC-32	Message Integrity Code	CBC-MAC	Secure Hash Algorithm
KEY MANAGEMENT	NA	4-Way handshaking mechanism	4-Way handshaking mechanism	Simultaneous Authentication
AUTHENTICATION	WPE-Open WPE-shared	Pre-Shared Key	Pre-Shared Key	Simultaneous Authentication of equals with EAP variant

C. Vulnerability/Weaknesses of WLAN Protocols

1. Vulnerability of WEP

As IV is short i.e of 24 bits, there can be cases when two packets are captured using the same IV. This shows the vulnerability of Short IV size. Also as there is no particular way to generate IV, there can be a possibility when wifi is using the same IV for a long period of time. This highlights the Keystream vulnerability. When plain text is associated with the unencrypted Integrity Check Value , it leads to brute force attacks by the attackers.

2. Vulnerability of WPA/WPA2-PSK

WPA is similar to WEP, but it uses temporal key integrity protocol(TKIP) to increase the encryption. TKIP encapsulates WEP using various algorithms. WPA2 on the other hand uses Advanced Encryption Standard. Communication of packets between the Client and Authenticator occurs using a 4-way handshake and it takes place between client and AP whenever a client tries to connect to an AP. PMK is calculated using the PBKDF2 hashing technique. By entering SSID, self-created

pass, and SSID length into this method, the attacker can build a hashed key and compare it to the captured hashed key. The AP and client verify that the credentials (WPA Key) used to initiate the connection are correct and then exchange the key to encrypt all the traffic from that point onwards.

The cryptographic key is installed whenever the client receives the third handshake message. A 4-way handshake's weak point lies at this stage, where messages might be misplaced or deleted. When the AP fails to receive message 4 (acknowledgement message) from the client, it re-transmits message 3 and lets the client receive it several times by reinstalling the same cryptographic key each time the third message is received. The assault takes place here.

3. Vulnerability of WPA3

WPA3 shows vulnerability in its DragonFly Handshake where dragonfly means the mechanism through which the user authenticates itself. It uses strong elliptical curves for encryption but hackers can force it to use weaker curves for encryption.Another vulnerability is side channel attack which uses an unprotected code, i.e which can be modified using if-then-else branch in Dragonfly algorithm to guess the password generation method.

D. Attacks on WLAN protocols

1. Attack on WEP

One can capture the packet and inject it into traffic to force an access point for creating a new packet with a new IV and continue to do so till there are two packets using the same IV. Using the vulnerability of short IV size one can figure out the secret key used in encrypting and decrypting the plain text.

1.1 Sniff all packets from target Access Point using the command - **"*airodump-ng --channel 2 --bssid <MAC ADDRESS> --write packets wlan0*".**

It will capture all packets and store them in .cap file format.

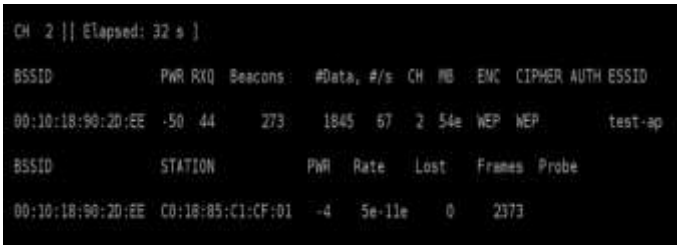


Figure 1. Sniff all the packets from the target AP

1.2 Crack WEP Key from the captured packets using the command - *"**aircrack-ng <filename of stored packets>**".*

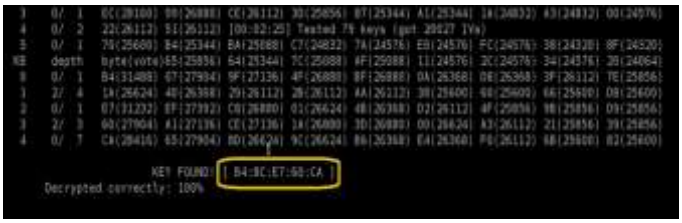


Figure 2. Cracking WEP key from the captured packets

2. Attack on WPA/WPA2

2.1 PTK is generated using PMK and PBKDF2 hashing function, this is the loophole for attackers to exploit as this data is sent in an unencrypted format. The pass, which can be computed via a dictionary attack with a solid 4-way handshake captured, is the attacker's lone unknown value in computing the PMK. We don't have a cryptographic flaw in WPA2, thus there's no other method to reverse engineer that key.

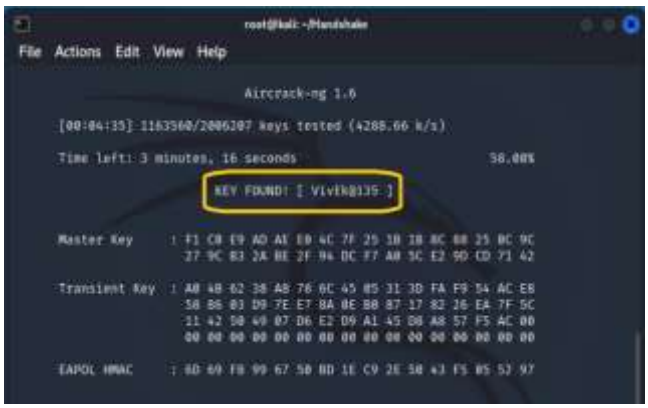


Fig 3. Cracking passphrase from the captured four-way handshake

2.3 KRACK for WPA2 (Key Reinstallation Attack)
It's used to get around the WPA2 protocol's weakness. As previously mentioned, the attacker can impersonate the AP by re-transmitting message-3 multiple times. When the client attempts to reconnect to the AP, the attacker can force it to connect to the phoney AP. It can operate as a middleman. Attackers can crack the pass with the captured handshake using brute force and dictionary assaults.

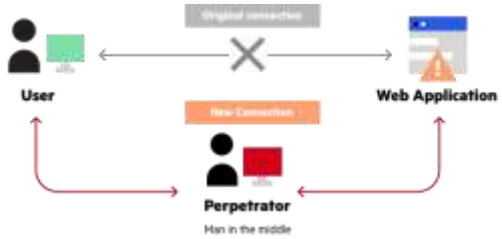


Fig 4. Cracking passphrase using KRACK

3. Attack on WPA3

Downgrade Attack - As some devices do not support new protocols, transition mode can be exploited through two different ways. First one is to modify the beacons by being man in the middle showing that a WPA3-enabled router can only be used as WPA2. And second is if the SSID name of the targeted WPA3 network is known, one can forge a man in the middle redirecting every request of WPA3 to connect to WPA2 access point. Once it act as WPA2 attacker can exploit the four way handshake of WPA2 above as explained in attacks of WPA2

E. MAC SPOOFING

For protecting the individual's privacy, mac address should be anonymous. **Therefore MAC SPOOFING is done before performing any kind of attack on the above mentioned protocols.**

Algorithm to spoof MAC address:

1 ifconfig wlan0 down (Down the interface)
2 macchanger -a wlan0 (Change the mac address)
3 ifconfig wlan0 up (Up the interface)

Store this code in the crontab so that whenever the system starts code execute itself and mac address get changed automatically.



Figure 5: Mac Spoofing Script

Pseudo-Code to spoof MAC address using macchanger.

```
Create a bash script using nano (filename).sh
Write the following command
    ifconfig (interface) down
    macchanger -A (interface)
    ifconfig (interface) up
Execute it using ./filename.sh
For changing MAC address every time you reboot, use the
crontab -e command as every time you reboot the system, a
command written in crontab gets executed.
Write the following in crontab:
    "@reboot /root/filename.sh"
```

Fig 6. Pseudo Code for MacChanger

VI. CONCLUSION

WEP was introduced in 1999 it has a vulnerability of short IV size which leads to cracking of WEP key. WPA was introduced in 2003 which uses TKIP that dynamically changes the key which system uses. In the case of WEP the key was static. Later on in 2004 WPA2 was introduced to mitigate the chances of brute force attacks which was seen in WPA by using Advanced Encryption Standard techniques. As key in WPA2 can be cracked using KRACK attack, later on in 2018 WPA3 protocol was introduced which has a different and longer key size as compared to the other protocols. Also WPA3 uses a simultaneous authentication method. Still some downgrade attacks can be performed on WPA3 as described above. Every protocol has some vulnerabilities which lead to cracking of password for different wifi. These are getting mitigated as soon as they are identified.

VII. REFERENCES

[1] G. Ola, "Penetration Testing on a Wireless Network .," 2013.

[2] A. O. Karen Scarfone, Murugiah Souppaya, Amanda Cody, "Technical Guide to Information Security Testing and Assessment Recommendations of the National Institute of Standards and Technology," *Nist Spec. Publ.*, vol. 800, pp. 1–80, 2008, [Online]. Available: https://nvlpubs.nist.gov/nistpubs/Legacy/SP/nistspecialpublication800.

[3] M. Vanhoef and F. Piessens, "Key Reinstallation Attacks," pp. 1313–1328, 2017, doi: 10.1145/3133956.3134027.

[4] N. Golmie, N. Chevrollier, and O. Rebala, "Bluetooth and WLAN coexistence: challenges and solutions," *IEEE Wirel. Commun.*, vol. 10, no. 6, pp. 22–29, Dec. 2003, doi: 10.1109/MWC.2003.1265849.

[5] M. Kyei and M. Asante, "Penetration Testing of IEEE 802.11 Encryption Protocols using Kali Linux Hacking Tools," *Int. J. Comput. Appl.*, vol. 176, no. 32, pp. 26–33, 2020, doi: 10.5120/ijca2020920365.

[6] T. Kropeit, "Don't Trust Open Hotspots: Wi-Fi Hacker Detection and Privacy Protection via Smartphone," 2015.

[7] A. Yacchirena, D. Alulema, D. Aguilar, D. Morocho, F. Encalada, and E. Granizo, "Analysis of attack and protection systems in Wi-Fi wireless networks under the Linux operating system," in *2016 IEEE International Conference on Automatica (ICA-ACCA)*, Oct. 2016, pp. 1–7, doi: 10.1109/ICA-ACCA.2016.7778423.

[8] H. Peng, "WIFI network information security analysis research," in *2012 2nd International Conference on Consumer Electronics, Communications and Networks (CECNet)*, Apr. 2012, pp. 2243–2245, doi: 10.1109/CECNet.2012.6201786.

[9] Y. Chen, W. Wang, and Q. Zhang, "Privacy-preserving location authentication in WiFi with fine-grained physical layer information," in *2014 IEEE Global Communications Conference*, Dec. 2014, pp. 4827–4832, doi: 10.1109/GLOCOM.2014.7037570.

[10] A. Ye, Q. Li, Q. Zhang, and B. Cheng, "Detection of Spoofing Attacks in WLAN-Based Positioning Systems Using WiFi Hotspot Tags," *IEEE Access*, vol. 8, no. 1, pp. 39768–39780, 2020, doi: 10.1109/ACCESS.2020.2976189.

[11] S. Dhall, S. K. Pal, and K. Sharma, "A chaos-based probabilistic block cipher for image encryption," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 1, pp. 1533–1543, Jan. 2022, doi: 10.1016/j.jksuci.2018.09.015.

[12] D. Kumar and D. S. S., "Enhancing Security Mechanisms for Healthcare Informatics Using Ubiquitous Cloud," *J. Ubiquitous Comput. Commun. Technol.*, vol. 2, no. 1, pp. 19–28, 2020, doi: 10.36548/jucct.2020.1.003.

[13] S. Shakya, "an Efficient Security Framework for Data Migration in a Cloud Computing Environment," *J. Artif. Intell. Capsul. Networks*, vol. 01, no. 01, pp. 45–53, 2019, doi: 10.36548/jaicn.2019.1.006.

[14] P. W. Nätverk, "PUBLIC WI-FI NETWORKS Sammanfattning."

[15] S. Sukhija and S. Gupta, "Wireless Network Security Protocols A Comparative Study," *Int. J. Emerg. Technol. Adv. Eng.*, vol. 2, no. 1, 2012.

AUTHORS

First Author – Sandesh Jain, Department of Information Technology, Delhi Technological University, Delhi,India, sandeshjain_2k18it109@dtu.ac.in

Second Author – Sarthak Pruthi, Department of Information Technology, Delhi Technological University, Delhi,India, sarthakpruthi_2k18it110@dtu.ac.in

Third Author – Vivek Yadav, Department of Information Technology, Delhi Technological University, Delhi,India, vivekyadav_2k18it131@dtu.ac.in

Correspondence Author – Dr. Kapil Sharma, Head of Department of Information Technology, Delhi Technological University, Delhi,India, kapil@ieee.org

Letters in Applied NanoBioScience
Open-Access Journal (ISSN: 2284-6808)

Article
Volume 12, Issue 3, 2023, 66
https://doi.org/10.33263/LIANBS123.066

Evaluation of Phytocompounds of *Vitis vinifera*, *Tabernaemontana corymbosa* and *Aloe vera* against JAB1: A Strategy to Treat Cervical Cancer

Ritika Luthra [1], Nikita Sharma [1], Shreeja Datta[1], Asmita Das [1,*]

[1] Department of Biotechnology, Delhi Technological University, Main Bawana Road, Delhi-110042
* Correspondence: asmitadas1710@dce.ac.in (A.D.);

Scopus Author ID 56416596400
Received: 4.02.2022; Accepted: 5.03.2022; Published: 15.04.2022

Abstract: JAB1 or C-Jun activation domain-binding protein-1 plays a role in controlling cellular proliferation, cell apoptosis, cell cycle, and regulating genomic instability as well as DNA damage response. JAB1 dysregulation contributes to oncogenesis via deactivating tumor suppressor genes. *Vitis vinifera, Tabernaemontana corymbosa,* and *Aloe vera* have been known for their antitumor functions. However, the inhibitory role of phytocompounds derived from *Vitis vinifera*, *Tabernaemontana corymbosa*, and *Aloe barbadensis miller* (Aloe vera) against JAB1 has not yet been elucidated. Therefore, we studied the naturally derived compounds from *Vitis vinifera*, *Tabernaemontana corymbosa*, and *Aloe vera*, which can interact and decrease the activities of JAB1 using in-silico analysis. Eighty-one compounds were screened using PyRx and further analyzed via Lipinski's rule. Further, on executing molecular docking studies on the sixty selected compounds, Iboluteine was found to be a potent inhibitor of JAB1 and should further be explored in *in-vitro* studies to substantiate these assertions. Phytocompounds are a reservoir of bioactive molecules that can be particularly useful in treating cervical cancer due to their limited toxicity and amenability from the development of drug resistance. Thus, this study aims to identify possible naturally derived compounds from *Vitis vinifera, Tabernaemontana corymbosa,* and *Aloe vera*, which can selectively decrease the activities of JAB1 that has the potential to be a crucial anti-cancer therapeutic molecule against cervical cancer.

Keywords: cervical cancer; JAB1; molecular docking; phytocompounds; tumor suppressor.

© 2022 by the authors. This article is an open-access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Cancer is amongst the biggest causes of mortality globally, with cervical cancer being the third most prevalent cancer among women. It is primarily caused by uncontrollable cell proliferation that results in a tumor [1,2]. Carcinogenesis in the cervix is caused by a variety of genetic as well as epigenetic changes. Even though HPV (Human papillomavirus) remains a major cause, several investigations show that the incidence is not a result of a single factor [3].

Jab1 or COPS5 (COP9 signalosome subunit 5) was found to be the 5th element of a constitutive photomorphogenic 9 signalosome (i.e., COPS5), a multi-purpose protein known to play a role in controlling a broad range of cell development mechanisms, such as cell signaling, growth, and differentiation, cell cycle, programmed cell death, DNA damage responses, but also tumorigenesis [4]. It has been regarded as an oncogene, which gets overexpressed abnormally in several cancers. An inverse relationship between p27 and Jab1 oncogene was found during the development of cervical cancer. Jab1 plays a pivotal role in

regulating the cell cycle by promoting the breakdown of several cyclin-dependent kinase inhibitors (CKIs) and cyclins. P27, known as a ubiquitous CKI, caused the inhibition of enzymatic activities of cyclin-cyclin dependent kinase complex, which resulted in a cell cycle halt at the G1 stage [5]. Jab1 represses p27 expression by facilitating the proteasome route, thus promoting cell growth. p27 deposition inside the nucleoplasm, halt in the cell cycle growth, and proliferation were observed when Jab1 was knocked out [6].

Furthermore, it was observed that Jab1 controls the release of Bcl-XL, cyclin-D, c-myc, p16, and p21, among other cell cycle regulators [7]. Besides, Jab1 can also act as a transcriptional co-factor for MYC that results in transcriptional activation of genes associated with processes like cell growth, angiogenesis, and invasion. It can also stabilize protein molecules like hypoxia-inducible factor-1α and c-Jun [8]. Such studies have shown that Jab1 interacts with the related cancer signaling mechanisms, indicating that it could be a prospective treatment target in cancers. Because Jab1 overexpression was linked to bad prognostics, the invention of the Jab1 selective inhibitor could have a big impact on cervical cancer treatment. However, to date, very few studies have given any evidence to the claim. Pandey [9] explained the potential of methanolic extract of *Moringa oleifera* against JAB1 in cervical cancer. The results showed enhanced ROS (Reactive oxygen species) production and caspase-3 activation, further explaining the probable mechanism behind the inhibitory effects of this plant's extract towards JAB1 in cervical cancer cells (particularly HeLa). As a result, this study proved the role of phytocompounds in targeting JAB1 in cervical cancer. Another study by Pandey et al. [10] established the relation of hesperidin with JAB1, which had the potential of preventing the progression of cervical cancer. The study further demonstrated that treatment using hesperidin ended up in the downregulation of JAB1 in a dose-dependent manner in the HeLa cell line, thus suggesting JAB1 as a promising therapeutic target for suppression of cell growth in HeLa cells. Pandey [11] also hypothesized that rutin exhibited inhibitory effects towards JAB1 and caused the considerable arrest of growth in SiHa cancer cells in a dose-dependent manner. This study further showed that rutin produced anti-cancer effects by modulating the expression of oncogenes and tumor suppressor genes.

Surgical intervention, particularly pelvic lymphadenectomy or radical hysterectomy, is the primarily opted treatment method for this cancer type. In addition, therapy with radiation therapy or chemotherapy is also resorted to. Unfortunately, each of these treatments is linked to serious adverse impacts, including bleeding, organ failure, and the danger of blood clots. Radiation therapies may cause menstrual irregularities, including menopause, irritation, discomfort during sexual activity, or even fertility issues. In contrast, chemotherapy can trigger cytotoxic effects throughout the entire body and not just in tumors [12]. Natural compounds have the potential to act as drugs or as a basis for the development of new drugs [13,14]. According to a recent study, over the last four decades, around 929 out of 1881 newly approved drugs have a natural origin, ranging from antibacterial, antitumor, antifungal, antiparasitic, antiviral, and so on [15]. Naturally, derived compounds are less toxic, and their extracts also show synergetic behavior [16]. Thereby, targeting cancer via these natural compounds is a useful alternative. The most economically effective treatment option for a variety of diseases possessing lesser adverse effects is making use of medicinal herbs [17, 18, 19]. Because of the antioxidant, anti-inflammatory, and antiproliferative properties of *Vitis vinifera* extracts, its phytochemicals can be used for nutraceutical formulations in treating and managing different cancers [20]. Different secondary metabolites derived from *Tabernaemontana* species (for instance, *Tabernaemontana corymbosa*) like terpenes, phenolics, lactones, flavonoids, steroids,

and alkaloids, have also been utilized as therapeutics due to their biological effects like anti-cancer, antimicrobial, analgesic, antioxidant, wound healing, anti-inflammatory, larvicidal, anticholinesterase, antivenom, antidiabetic and antihypertensive action [21]. In addition to this, another plant, *Aloe vera*, has properties like growth suppressor evasion, prevention of immune destruction, tumor-promoting inflammation, activation of invasion and metastasis, induction of angiogenesis, genome instability and mutation, cell death resistance, deregulation of cellular energetics, as well as sustainment of proliferating signaling, thus proving its anti-cancer activities [22]. For this reason, these three plants have been considered in the present study to find potential natural inhibitors of Jab1 using computational techniques.

2. Materials and Methods

2.1. Tools required for in silico studies.

The software and databases used in this study are IMPPAT (Indian Medicinal Plants, Phytochemistry And Therapeutics) database, Protein Data Bank, PubChem, Alpha fold protein structure database, Discovery studio visualizer Biovia, PyMOL, SwissADME, PyRx, and AutoDock v4.2.6. 2.2.

2.2. Ligand selection.

In total, eighty-one ligands were selected from the phytochemical constituents of three different plants: *Vitis vinifera, Tabernaemontana corymbosa*, and *Aloe Vera* from the literature [23-25] and database called IMPPAT:- https://cb.imsc.res.in/imppat [26]. The base of selection of ligands majorly depends upon the phytocompounds showing anti-cancer properties. The structures of these ligands molecules were procured from a database known as Pubchem [27]. The ligands (Refer to Table 1) were downloaded into 3D structures in .sdf format. These downloaded ligands structures were further converted into .pdb format through Biovia discovery studio's visualizer to run molecular docking through Autodock v4.2.6.

2.3. Receptor molecule.

The three-dimensional structure of Jab1/CSN5 protein (with Uniprot ID: Q92905 and https://alphafold.ebi.ac.uk/entry/Q92905), which has a major role in the progression of cervical cancer, was obtained from AlphaFold Protein structure database (https://alphafold.ebi.ac.uk/), an AI-based online system used for predicting 3D structures of a protein via its amino acid sequence. The structure was downloaded in .pdb format.

2.4. Virtual screening of ligands.

The screening of phytocompounds of all three plants was done through virtual screenings software: - PyRx [28]. This software was used to screen out the ligands which were showing minimum binding energy with the targeted protein (Refer to Table 1). The phytocompounds showing minimum binding affinity were further investigated for ADME analysis. The protein and ligands present in PyRx run on .pdbqt format. The procedure of PyRx started with loading the macromolecule (protein) present in .pdb format. The macromolecule was further converted into .pdbqt format. After the conversion of the macromolecule, ligands that were converted into .pdb format were loaded in PyRx. The energy of all ligands was minimized, followed by the conversion of ligands into .pdbqt format. The docking of all ligands

and the macromolecule was performed by forming a grid, and molecules having lower binding energy were screened out for further analysis. The resultant structures of the docked complex were saved into .pdb format.

2.5. ADME analysis.

Drug likeness property analysis or ADME analysis was done through a tool known as SwissADME (https://www.swissadme.ch/). It is an online tool used to do the screening of compounds that follow the five Lipinski rules of drugs [29]. The compounds violating the rules were eliminated, and leftover ligands were used for molecular docking. Here, the SMILE notations of screened ligands were taken from PubChem and were pasted to the online server of SwissADME. The five Lipinski rules state that: - The H-bond donor should be less than 5, the molecular weight of the compound should be less than 500Da, the H-bond acceptor should be less than 10, lipophilicity (LogP) should be less than 5, and molar refractivity should be in between 40 to 130. The phytocompounds that violated more than one of Lipinski's rules were not considered for molecular docking [30].

2.6. Protein preparation.

Before carrying out the docking process, the Jab1/CSN5 protein was prepared by removing the water molecules. This was followed by removing Hetatoms and saving the file in .pdb format to perform further analysis.

2.7. Molecular docking.

The docking experiments of ligands against Jab1 protein were performed through Autodock v4.2.6. This software was used to demonstrate the binding energy analysis via energy potential and grid formation. Different search algorithms were used to find the specific binding energy features on the targeted protein [31]. In the AutoDock, water molecules were removed from the prepared macromolecule and shortlisted ligands. In contrast, polar hydrogen molecules and charges like Gasteiger and Kollman were added to the Jab1 protein, summing up the optimization of ligand and targeted protein. The loaded ligands and protein were saved into .pdbqt format. The active site residues of Jab1 protein were retrieved from a protein visualization software, PyMOL. These active site residues were taken to form a grid box of 85x85x85 with a spacing of 0.375Å. The .pdbqt format files of both ligand and targeted protein were used to run autogrid. The genetic algorithm was kept as the search parameter and the docking research as the output which provided the information on Lamarckian GA. The docking log files were acquired for further binding energy analysis. This procedure was repeated three times for every loaded ligand, and the final conformation possessing the highest negative binding affinity among the ten energy conformations was taken into account. The docked structure was then converted into a two-dimensional structure used to analyze the interactions formed at the binding site of Jab1 protein and ligand through Biovia Discovery Studio Visualiser.

2.8. Bioavailability radar testing.

A more reasonable evaluation of physicochemical properties was used to extend the search for some potent ligand molecules. Bioavailability radars against the ligands that outperformed the standard compound were created using the SwissADME (an internet tool).

Six decision variables were used to analyze the compounds: flexibility, lipophilicity, polarity, saturation, size, and solubility. Molecules that varied from either of the standard values anticipated being orally inaccessible, so they were removed from a further study [29].

2.9. PASS webserver prediction.

The PASS webserver is an online server that can be used to predict the biological activity of a ligand or a compound. The SMILES of ligand were taken from PubChem as an input, and it provides the probability of biological activity. It tells the information about a drug's activeness or inactiveness in that particular biological activity.

3. Results and Discussion

3.1. Virtual screening of ligands.

The ligands subjected to virtual screening were shortlisted for ADME analysis on the basis of their binding energy. Eighty-one ligands underwent virtual screening, and the ligands showing binding energy of -3.5 Kcal/mol and above were selected for ADME analysis. Out of eighty-one ligands, only seventy-one had binding energy of -3.5 Kcal/mol or more than that. The remaining ten ligands were not accounted for ADME analysis (Table 1).

Table 1. Results of virtual screening of phytocompounds via PyRx.

Plant Name	Phytocompound Name	Binding affinity (Kcal/mol)
Vitis vinifera	(-)-Epicatechin gallate	-8.7
	Ampelopsin A	-8.6
	29-Nor-20-oxolupeol	-8.4
	Luoenone	-8
	Betulin	-7.9
	(+)-Epicatechin	-7.5
	Delphinidin	-7.4
	Dihydromyricetin	-7.4
	Resveratol	-7.4
	Pterostilbene	-6.8
	2,4-Dihydroxycinnamic acid	-6.2
	Alpha-terpineol	-5.6
	8-hydroxygeraniol	-5.3
	(-)-Linalool	-5.1
	Dehydroascorbic Acid	-5
	2,6-Dimethyloct-7-ene-2,3,6-triol	-4.9
	2(3H)-Furanone, 5-ethoxydihydro-	-4.6
	Eupatin	-4.7
	Malvidin	-4.7
	3-beta-Hydroxy-20(29)-lupene	-4.6
	Brevilagin 1	-4.6
	Alpha-Viniferin	-4.5
	Astilbin	-4.4
	Carotene	-4.3
	1,3-Propanediol	-3.5
	Vitisin C	-3.2
	Pseudomonic acid A	-3.1

Plant Name	Phytocompound Name	Binding affinity (Kcal/mol)
Tabernaemontana corymbosa	Ervachinine B	-9
	Ervachinine D	-8.7
	Velbanamine	-8.6
	Criofolinine	-8.5
	Jerantinine C	-8.3
	Lirofolines A	-8.2
	Affinisine	-8
	16-epivobasine	-7.9
	Iboluteine	-7.9
	Jerantinine E	-7.8
	Conolutinine	-7.7
	Jerantinine F	-7.7
	Vobasenal	-7.7
	Ervaluteine	-7.6
	Vernavosine	-7.5
	19-epi-heyneanine	-7.4
	Jerantinine B	-7.4
	Iboxygaine	-7.3
	Jerantinine A	-7.1
	Jerantinine D	-7
	Conodiparine A	-6.4
	Conodiparine B	-6.3
	Vincristine	-6.2
	Heyneanine	-3.4
	Tabernaricatine C	-3.3
	Lirofolines B	-3.2
Aloe Vera	(E)-Cycloarta-23-ene-3beta-ol	-8.5
	Chrysophanol	-8.1
	10-Hydroxyaloin A	-8
	Apigenin	-8
	Lophenol	-8
	Aloesaponarin I	-7.9
	Aloin	-7.9
	Barbaloin	-7.8
	Emodin	-7.8
	Aloe-emodin	-7.7
	Aloesaponarin II	-7.7
	Aloesone	-7.1
	Aloesin	-6.8
	Esculetin	-6.4
	3,5-Di-tert-butyl-4-hydroxybenzaldehyde	-6.3
	Umbelliferone	-6.2
	Benzothiazolone	-5.4
	Allantoin	-5.2
	15-Methylhexadecanoic acid	-4.5
	7-Hydroxyaloin B	-4.4
	Trypan Blue free acid	-3.8
	Octapeptide	-3.7

Plant Name	Phytocompound Name	Binding affinity (Kcal/mol)
	Beta-carotene	-3.6
	Galactitol	-3.4
	CHEMBL518845	-3.4
	Creatinine	-3.3
	Salicylic acid	-3.2
	6-Biopterin	-2.8

3.2. In-silico bioavailability analysis.

All seventy-one ligands were analyzed using the SwissADME tool on the Swiss Institute of Bioinformatics page (refer to http:/www.swissadme.ch/) for virtual Physico-chemical features like drug resemblance, water-solubility, pharmacokinetics, lipophilicity, medicinal chemistry, and pharmacokinetics parameters [29].

3.3. ADME Analysis.

The shortlisted seventy-one phytocompounds derived from the three selected plants were further subjected to preliminary comparisons (Table 2), conducted with respect to ligand molecule properties. These comparisons helped rule out the compounds that did not satisfy the properties of becoming a drug candidate. The five rules of Lipinski were used on seventy-one phytocompounds for estimating the drug-likeness, and the phytocompounds which obeyed the Lipinski rules were selected for further investigations. Compounds that did not follow two or more Lipinski rules were not pursued further. The remaining phytocompounds were considered for molecular docking studies. Eleven compounds, namely, 3beta-Hydroxy-20(29)-lupene, brevilagin 1, alpha-Viniferin, astilbin, carotene (derived from *Vitis vinifera*), conodiparine A, conodiparine B, vincristine (derived from *Tabernaeumontana corymbosa*), acemannan, trypan blue free acid, octapeptide (derived from *Aloe vera*) violated more than one Lipinski rule. Therefore, the remaining 60 compounds were examined for docking studies.

Table 2. ADME analysis of selected phyto-compounds.

Plant Name	Phytocompound	Molecular weight	H-bond acceptors	H-bond donors	Molar Refractivity	Lipophilicity (LogP < 5)	Lipinski Violation
Vitis vinifera	29-Nor-20-oxolupeol	428.69 g/mol	2	1	131.01	7.04	1
	3beta-Hydroxy-20(29)-lupene	426.72 g/mol	12	1	135.14	9.87	2
	Brevilagin 1	354.22 g/mol	9	7	-	-1.76	2
	Lupenone	424.70 g/mol	1	0	134.18	9.56	1
	(-)-Epicatechin gallate	442.37 g/mol	10	7	110.04	1.53	1
	(-)-Linalool	154.25 g/mol	1	1	50.44	2.97	0
	(+)-Epicatechin	290.27 g/mol	6	5	74.33	0.36	0
	1,3-Propanediol	76.09 g/mol	2	2	18.86	-1.04	0
	2,4-Dihydroxycinnamic acid	180.16 g/mol	4	3	47.16	1.16	0
	2,6-Dimethyloct-7-ene-2,3,6-triol	188.26 g/mol	3	3	53.27	0.44	0
	5-Ethoxydihydro-2(3H)-furanone	130.14 g/mol	3	0	31.21	0.56	0
	8-hydroxygeraniol	170.25 g/mol	2	2	51.56	2.31	0
	alpha-Terpineol	154.25 g/mol	1	1	48.8	3.39	0
	alpha-Viniferin	662.68 g/mol	9	6	185.07	6.83	3
	Ampelopsin A	470.47 g/mol	7	6	128.34	3.77	1
	Astilbin	450.40 g/mol	11	7	-	0.04	2
	Betulin	442.72 g/mol	2	2	136.3	8.28	1

Plant Name	Phytocompound	Molecular weight	H-bond acceptors	H-bond donors	Molar Refractivity	Lipophilicity (LogP < 5)	Lipinski Violation
	Carotene	536.89 g/mol	0	0	-	12.46	2
	Dehydroascorbic Acid	174.11 g/mol	6	2	32.85	-1.03	0
	Delphinidin	303.24 g/mol	7	6	78.2	1.22	1
	Dihydromyricetin	320.25 g/mol	8	6	76.78	0.59	1
	Resveratol	228.24 g/mol	3	3	67.88	3.13	0
	Pterostilbene	256.30 g/mol	3	1	76.82	3.78	0
	Eupatin	360.31 g/mol	8	3	93.47	2.89	0
	Malvidin	331.30 g/mol	7	4	87.13	2.24	0
Tabernaemontana corymbosa	Jerantinine A	382.45 g/mol	5	2	112.97	2.63	0
	Velbanamine	382.45 g/mol	5	1	110.39	2.85	0
	Jerantinine B	398.45 g/mol	6	2	112.41	1.79	0
	16-epivobasine	402.91 g/mol	4	0	116.79	2.93	0
	Jerantinine C	396.44 g/mol	5	2	113.17	1.97	0
	Conodiparine A	750.92 g/mol	9	4	218.65	4.44	2
	Iboxygaine	326.43 g/mol	3	2	99.62	2.34	0
	Jerantinine D	412.44 g/mol	6	2	112.61	1.25	0
	Iboluteine	326.43 g/mol	3	1	101.83	3.12	0
	Conodiparine B	750.92 g/mol	9	4	218.65	4.42	2
	Jerantinine E	384.47 g/mol	5	2	113.44	2.74	0
	Vincristine	824.96 g/mol	12	3	233.11	3.39	2
	Affinisine	308.42 g/mol	2	1	97.55	2.04	0
	Ervachinine D	676.89 g/mol	6	3	206.71	5.61	1
	Lirofolines A	648.83 g/mol	6	0	197.94	6.2	1
	Criofolinine	398.41 g/mol	6	3	105.84	1.27	0
	Conolutinine	312.41 g/mol	3	1	95.07	2.38	0
	19-epi-heyneanine	296.41 g/mol	2	2	93.12	2.8	0
	Ervachinine B	676.89 g/mol	6	3	206.71	5.61	1
	Jerantinine F	398.45 g/mol	6	2	112.41	1.9	0
	Vobasenal	352.38 g/mol	4	1	100.32	1.75	0
	Ervaluteine	342.43 g/mol	4	2	102.99	2.5	0
	Vernavosine	528.38 g/mol	5	2	128.54	3.03	1
Aloe Vera	Aloe-emodin	270.24 g/mol	5	3	69.92	1.82	0
	Aloin	418.39 g/mol	9	7	101.96	-0.12	1
	Chrysophanol	254.24 g/mol	4	2	68.76	3.53	0
	Aloesaponarin I	312.27 g/mol	6	2	80.04	3.48	0
	Aloesaponarin II	254.24 g/mol	4	2	68.76	3.07	0
	Aloesin	394.37 g/mol	9	5	96.87	-1.28	0
	Umbelliferone	162.14 g/mol	3	1	44.51	1.58	0
	Trypan Blue free acid	872.89 g/mol	16	8	212.77	4.35	3
	Esculetin	178.14 g/mol	4	2	46.53	1.22	0
	Allantoin	158.12 g/mol	3	4	39.55	-2.17	0
	Emodin	270.24 g/mol	5	3	70.78	2.72	0
	Octapeptide	1081.31 g/mol	12	12	291.73	0.7	3
	Aloesone	232.24 g/mol	4	1	64.25	1.18	0
	Apigenin	270.24 g/mol	5	3	73.99	3.02	0
	(E)-Cycloarta-23-ene-3beta-ol	426.72 g/mol	1	1	135.14	9.37	1
	10-Hydroxyaloin A	434.39 g/mol	10	8	103.01	-1.21	1

Plant Name	Phytocompound	Molecular weight	H-bond acceptors	H-bond donors	Molar Refractivity	Lipophilicity (LogP < 5)	Lipinski Violation
	15-Methylhexadecanoic acid	270.45 g/mol	2	1	85.6	7.42	1
	7-Hydroxyaloin B	434.39 g/mol	10	8	103.98	-1.19	1
	3,5-Di-tert-butyl-4-hydroxybenzaldehyde	234.33 g/mol	2	1	72.39	4.37	0
	Barbaloin	418.39 g/mol	9	7	101.96	-0.12	1
	Benzothiazolone	151.19 g/mol	1	1	42.45	1.76	0
	Beta-carotene	536.87 g/mol	0	0	184.43	13.54	2
	Lophenol	400.68 g/mol	1	1	128.42	8.72	1

3.4.1. ADME test.

The development of any drug requires analysis of ADME properties, namely, excretion, absorption, metabolism, and distribution, in the discovery process. Iboluteine was assessed for Swiss ADME. Distinctive to Swiss ADME, it depicts the graphical representation of the drug-likeness specifications of an oral bioactive drug. This graph is outlined as a hexagon in Fig. 1. The pink area present in the hexagon represents the ideal array for each property (lipophilicity, that is, XLOGP3 was 3.48; molecular weight was 326.43 g/mol; polarity, that is, TPSA was 41.57 Å^2; solubility, that is log s was 4; saturation value was 0.65 and carbon portions in the sp3 hybridization, as well as 2 rotatable bonds discovered flexibility) (Figure 1).

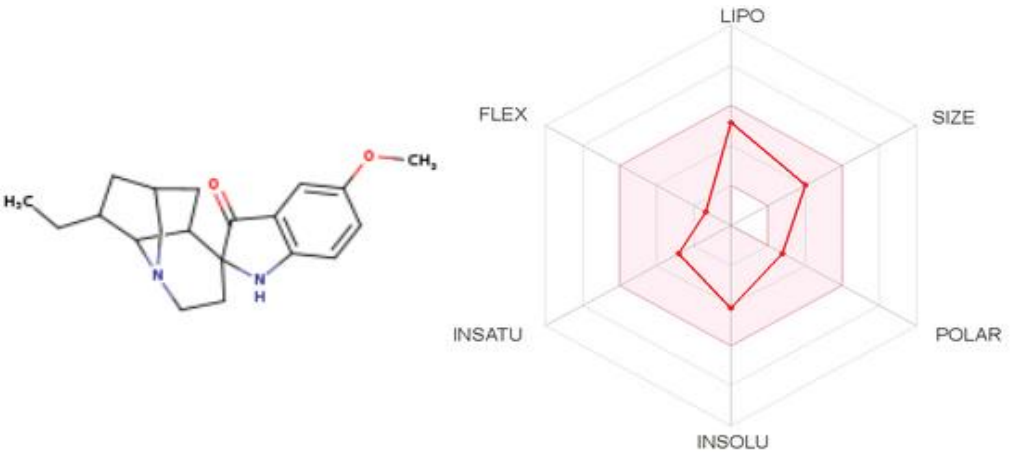


Figure 1. Structural features and bioavailability radar of ligand Iboluteine by SwissADME.

3.4.2. Drug-likeness.

Drug-likeness properties of Iboluteine are described by the distorted red hexagon inside the pink color. The molecule lies within the drug-likeness specifications of bioavailable drugs, further following the Lipinski rule with a bioavailability score of 0.55. The SwissADME even comprises computational filters, making the Lipinski rule the most common specification for evaluating small molecules for drug-likeness (Figure 2A).

3.4.3. Medicinal chemistry evaluation.

Evaluation of medicinal chemistry is used to predict drug-likeness of small molecules, including PAINS (Pan-assay interference compounds) and Brenk screening rules. The PAINS computer screening model is used to identify compounds that look in various high-performance biochemical screens as hits (or promiscuous compounds). Our molecule does not display any such PAINS alertness. In another selection model, this molecule followed the Brenk screening

law with only an isolated alkene, which determines that the compound is smaller as well as less hydrophobic and not those defined by Lipinski's rule (Figure 2A).

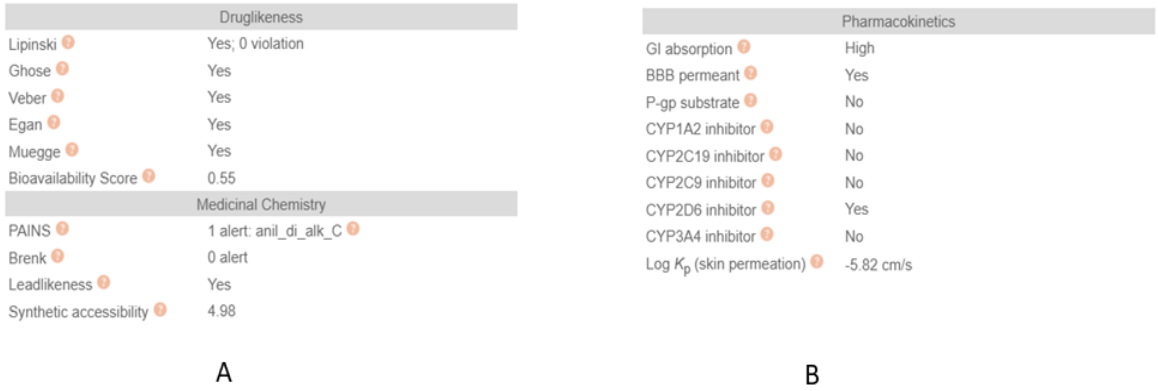


Figure 2. (**A**) Evaluation of medicinal drug likeliness and therapeutic chemistry of ligand Iboluteine, (**B**) Bioavailability radar of ligand Iboluteine (in the left panel) and pharmacokinetics (in the right panel) analyzed using SwissADME

3.4.4. Pharmacokinetics.

Using SwissADME, it was even possible to evaluate whether a chemical is a p-glycoprotein (P-gp) substrate or CYP (Cytochrome P450) isoenzyme inhibitor. The drug metabolism via CYP isoenzymes is a crucial cause of drug contacts that might end up in drug harmfulness and decreased pharmacological results. The model reviews if the molecule is more probable to be a substrate or not of a specified CYPP-gp substrate or an inhibitor or non-inhibitor. By inhibiting the CYP2D6 (Cytochrome P450 2D6) inhibitor, our ligand displayed pharmacological properties. This molecule depicted low Gi absorption and Log Kp (skin permeation) and falling in a negative range of -5.82 cm/s. Ligand crosses the blood-brain barrier. Further, *in-vitro* experiments will be needed to evaluate if Iboluteine is activated or deactivated by CYP2C19 (Cytochrome P450 2C19), CYP2C9 (Cytochrome P450 family 2 subfamily C member 9), as well as CYP3A4 (Cytochrome P450 3A4) (Figure 2B).

3.5. Molecular docking.

Molecular docking is an important computational analysis step that is used in drug design. Docking studies can predict the best orientation of the selected ligand with a targeted protein and helps in the production of conformation of the docked complex with the minimum binding affinity. All the sixty ligands which were obeying the five rules of Lipinski were subjected to molecular docking as per the active sites of Jab1 protein. The binding energies of standard drugs [32], which are used in cervical cancer against Jab1 protein, namely Irinotecan, Curcumin, Doxorubicin, and 5 Fluorouracil, were showing binding energies of -7.37 Kcal/mol, -5.48 Kcal/mol, -4.77 Kcal/mol, and -4.37 Kcal/mol respectively (Table 3).

Table 3. Molecular docking studies of selected phyto-compounds.

Plant Name	Phytocompound	Binding Energy (ΔG) (Kcal/mol)	Ligand Efficiency	Inhibition constants (μM)	Intermolecular energy	Vdw H bond desolvation energy
Vitis vinifera	Lupenone	-8.26	-0.27	0.62414	-8.76	-8.72
	29-Nor-20-oxolupeol	-7.7	-0.25	2.28	-8.29	-8.14
	Betulin	-7.22	-0.23	5.11	-8.41	-8.33
	Ampelopsin A	-6.87	-0.2	9.17	-9.26	-9.02
	Resveratol	-6.53	-0.38	16.31	-8.02	-7.69

Plant Name	Phytocompound	Binding Energy (ΔG) (Kcal/mol)	Ligand Efficiency	Inhibition constants (μM)	Intermolecular energy	Vdw H bond desolvation energy
	(+)-Epicatechin	-6.27	-0.3	25.51	-8.06	-7.86
	Malvidin	-6	-0.25	40.06	-8.09	-7.84
	Eupatin	-5.55	-0.21	85.59	-7.64	-7.37
	Pterostilbene	-5.39	-0.28	111.28	-6.89	-6.67
	Delphinidin	-5.28	-0.24	134.14	-7.37	-7.11
	Dihydromyricetin	-5.12	-0.22	176.2	-7.21	-7.09
	(-)-Epicatechin gallate	-4.91	-0.15	252.6	-8.19	-8.02
	alpha-TERPINEOL	-4.87	-0.44	271.35	-5.46	-5.39
	2(3H)-Furanone, 5-ethoxydihydro-	-4.52	-0.5	487.99	-5.11	-4.97
	2,4-Dihydroxycinnamic acid	-4.46	-0.34	534.75	-5.66	-5.66
	Dehydroascorbic acid	-4.37	-0.36	626.76	-5.56	-5.06
	(-)-Linalool	-4.26	-0.39	747.86	-5.76	-5.64
	8-hydroxygeraniol	-3.84	-0.32	1540	-5.92	-5.79
	2,6-Dimethyloct-7-ene-2,3,6-triol	-3.56	-0.27	2470	-5.94	-5.76
	1,3-propanediol	-2.57	-0.51	13170	-3.76	-3.28
Tabernaemontana corymbosa	Iboluteine	-8.33	-0.33	1.5	-8.54	-8.31
	Ervaluteine	-7.89	-0.31	1.96	-8.68	-8.41
	Ervachinine D	-7.59	-0.15	-2.73	-9.68	-9.74
	19-epi-heyneanine	-7.55	-0.34	2.91	-8.15	-7.8
	Ervachinine B	-7.46	-0.15	3.42	-9.54	-9.59
	Lirofolines A	-7.32	-0.31	4.32	-8.21	-7.9
	Vobasenal	-7.01	-0.27	7.25	-7.91	-7.52
	16-epivobasine	-6.83	-0.25	-9.92	-7.42	-7.11
	Iboxygaine	-6.74	-0.28	11.4	-7.64	-7.19
	Velbanamine	-6.73	-0.24	-11.71	-7.62	-7.05
	Vernavosine	-6.6	-0.23	14.59	-7.79	-7.2
	Affinisine	-6.55	-0.28	17.32	-7.09	-6.75
	Jerantinine D	-6.5	-0.22	17.07	-8	-7.79
	Conolutinine	-6.44	-0.28	19.09	-7.03	-6.58
	Jerantinine B	-6.38	-0.22	20.98	-7.87	-7.53
	Criofolinine	-6.34	-0.22	22.58	-7.53	-7.47
	Jerantinine F	-6.2	-0.21	28.33	-7.4	-6.5
	Jerantinine C	-5.88	-0.2	48.88	-7.37	-7.03
	Jerantinine A	-5.6	-0.2	78.04	-7.1	-7.11
	Jerantinine E	-5.52	-0.2	90.05	-7.01	-6.67
Aloe Vera	(E)-Cycloarta-23-ene-3beta-ol	-7.99	-0.26	1.38	-9.49	-949
	Lophenol	-7.56	-0.26	2.85	-9.35	-9.33
	Aloesin	-6.88	-0.25	9	-9.57	-9.09
	apigenin	-6.72	-0.34	11.96	-7.91	-7.58
	Aloesone	-6.49	-0.38	17.47	-7.39	-7.27
	Aloin	-6.35	-0.21	22.25	-9.33	-8.84
	Aloesaponarin II	-6.33	-0.33	22.73	-6.93	-6.77
	Umbelliferone	-6.06	-0.51	36.44	-6.35	-6.18
	Emodin	-5.96	-0.3	42.61	-6.86	-6.49
	Barbaloin	-5.88	-0.2	48.99	-8.86	-8.73
	Aloesaponarin I	-5.83	-0.25	53.24	-7.02	-6.82
	Aloe-emodin	-5.64	-0.28	73.17	-6.84	-6.66

Plant Name	Phytocompound	Binding Energy (ΔG) (Kcal/mol)	Ligand Efficiency	Inhibition constants (μM)	Intermolecular energy	Vdw H bond desolvation energy
	Esculetin	-5.6	-0.43	79.03	-6.19	-6.08
	Chrysophanol	-5.57	-0.29	82.9	-6.16	-6.05
	10-Hydroxyaloin A	-5.35	-0.17	119.11	-8.63	-8.52
	3,5-Di-tert-butyl-4-hydroxybenzaldehyde	-5.35	-0.31	119.48	-6.54	-6.51
	7-Hydroxyaloin B	-5.11	-0.16	178.57	-8.39	-8.32
	Allantoin	-4.98	-0.45	223.91	-5.28	-5.19
	15-Methylhexadecanoic acid	-4.86	-0.26	273	-9.34	-7.49
	Benzothiazolone	-4.59	-0.46	432.33	-4.59	-4.59
	Irinotecan (Standard)	-7.37	-0.17	3.93	-9.16	-9.2
	Curcumin (Standard)	-5.48	-0.2	96.08	-8.46	-8.52
	Doxorubicin (Standard)	-4.77	-0.12	317.81	-8.05	-8.1
	5-fluorouracil (Standard)	-4.37	-0.49	-627.09	-4.37	-4.34

The minimum binding energy among all the complexes of *Vitis vinifera*, *Tabernaemontana corymbosa*, and *Aloe vera* were -8.26 Kcal/mol, -8.33 Kcal/mol, and -7.99 Kcal/mol, respectively.

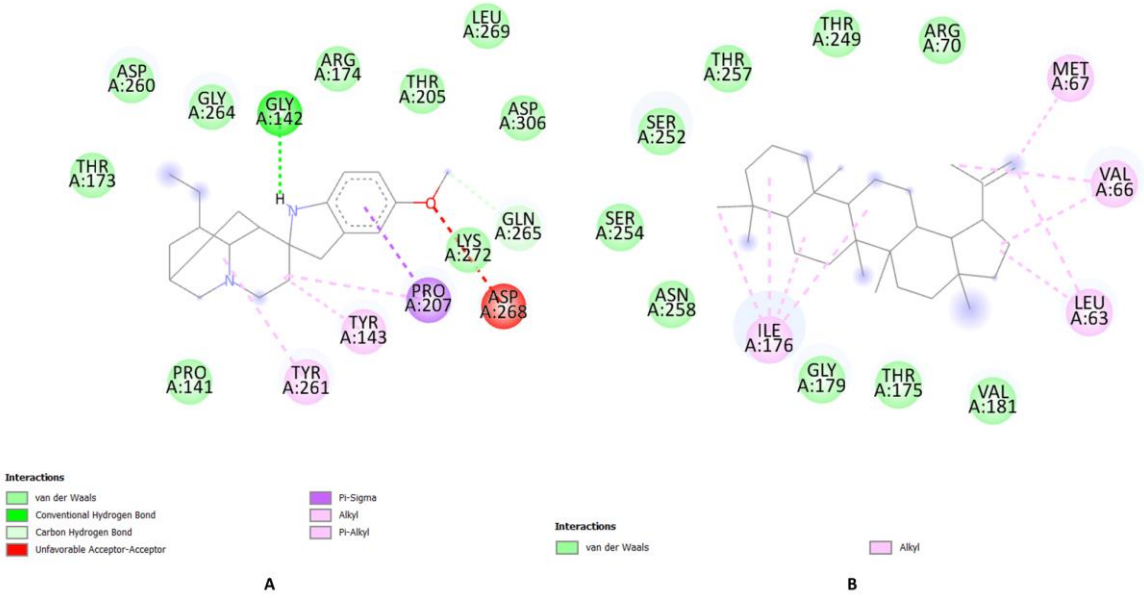


Figure 3. (**A**) Interaction between Iboluteine and Jab1; (**B**) Interaction between Lupenone and Jab1.

The minimum binding energy among all the selected ligands was Iboluteine (-8.33 Kcal/mol), a phytocompound of *Tabernaemontana corymbosa*. A total of fifteen residues interacted with ligand molecules. Only one residue, namely GLY142, was seen to form the strong conventional H-bond with the ligand. TYR261, TYR143, and PRO207 were seen to form another type of interaction, namely the Pi-Alkyl bond. PRO207 was also seen to form a Pi-sigma interaction. ASP268 showed unfavorable Acceptor-Acceptor interaction while GLN265 was forming a carbon-hydrogen bond with the ligand. The remaining residues interacted with van der Waals formation (Figure 3A) (Table 3).

Figure 3B represents the best docked binding energy conformation of the Lupenone complex. The binding energy was -8.26 Kcal/mol. Thirteen residues showed interaction with the ligand. Only two types of interaction were seen, Van der Waals interaction and Alkyl bond

formation. ILE176, LEU63, VAL66, and MET67 showed alkyl bond formation at different positions present in the ligand. The remaining residues ARG70, THR249, THR257, SER252, SER254, ASN258, GLY179, THR175, and VAL181 showed Van der Waals interaction with ligand (Table 3).

Among the ten energy conformations obtained for (E)-Cycloarta-23-ene-3beta-ol complex, -7.99 Kcal/mol was the best binding energy conformation. In total, twelve amino acid residues interacted with the given ligand. Out of twelve, five residues: LEU241, LEU238, LEU63, LEU62, and VAL66, interacted with the ligand by forming an Alkyl bond. TRP242 formed a Pi-sigma bond. The remaining residues like TRP246, TYR245, ASN248, THR249 VAL181, and ALA59 formed van der Waals bonds, as shown in Figure 4A and Table 3.

Ervaluteine, a phytocompound of *Tabernaemontana corymbosa*, showed the second-highest binding energy among all the phytocompounds of the same plant and held the fourth position with respect to binding energy among all sixty phytocompounds (Table 3). A total of seventeen residues interacted with the given ligand. GLY142 formed a strong conventional Hydrogen bond while PRO207 interacted via a Pi-sigma bond with the ligand. ILE 206 showed Amide-Pi stacked bond formation, and TYR261 was interacting through Pi-Alkyl bond formation. The remaining eleven residues showed weak van der Waals bond formation. Only GLN265 was seen to form a carbon Hydrogen bond (Figure 4 (B)).

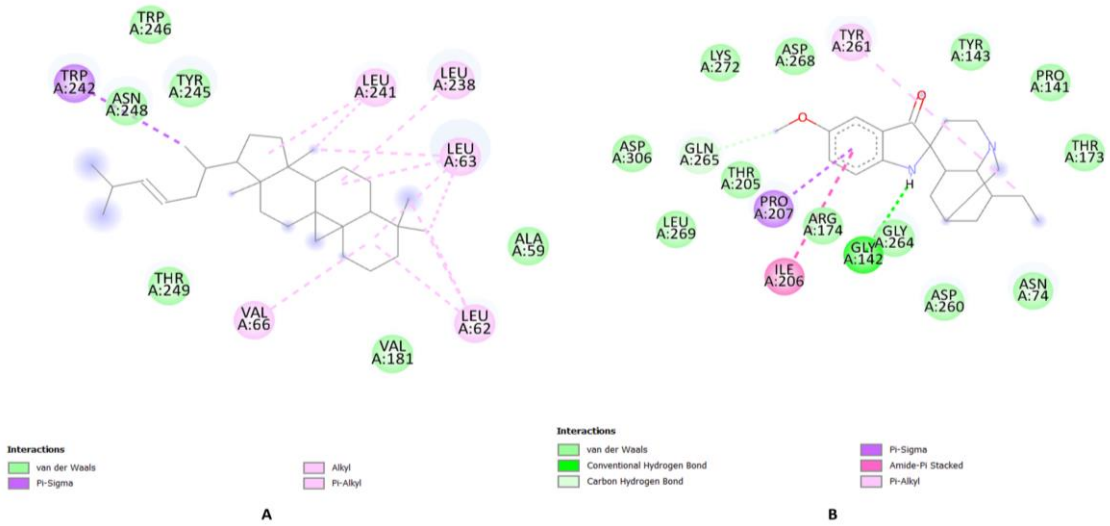


Figure 4. (**A**) Interaction between (E)-Cycloarta-23-ene-3beta-ol and Jab1; (**B**) Interaction between Ervaluteine and Jab1.

3.6. BioavailabilityrRadar.

Iboluteine and Ervaluteine of *Tabernaeumontana corymbosa*, Lupenone of *Vitis vinifera*, and (E)-Cycloarta-23-ene-3beta-ol of *Aloe vera* were among the top three docked compounds that were studied subsequently. Bioavailability radar is employed to conduct a much more detailed and complete analysis. This is a describing method that uses six physical and chemical parameters to assess the drug-likeness of selected compounds. In the case of Iboluteine and Ervaluteine, the ligand spectrum completely fitted in the pink-colored region, indicating that it is orally bioavailable. Both, Lupenone and (E)-Cycloarta-23-ene-3beta-ol, on the other hand, did not obey the solubility and lipophilicity parameters, concluding that these two compounds are not orally bioavailable (Figure 5).

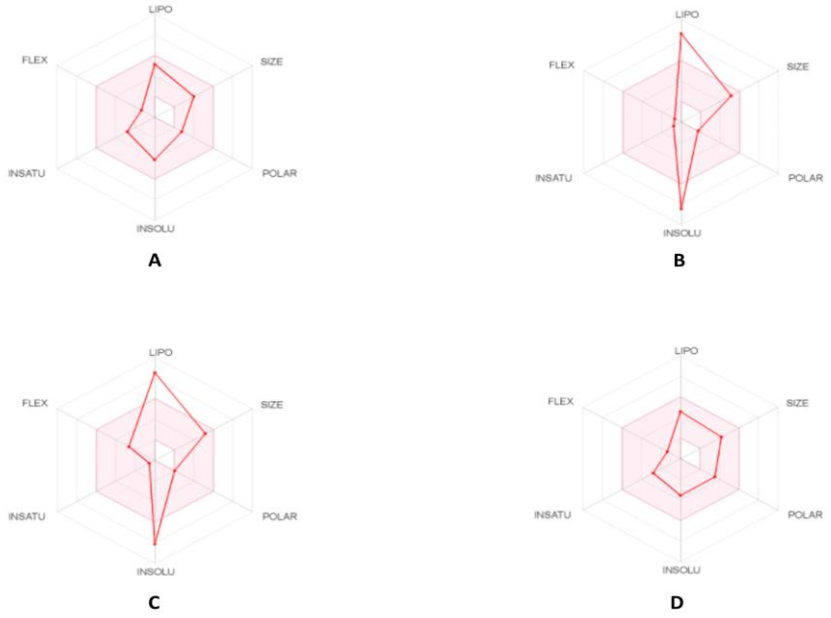


Figure 5. Bioavailability radars of (**A**) Iboluteine; (**B**) Lupenone; (**C**) (E)-Cycloarta-23-ene-3beta-ol; (**D**) Ervaluteine.

3.7. PASS Webserver Prediction.

The PASS webserver was used to predict the biological activity of a ligand-based on its conformation. The shortlist three compounds of *Vitis vinifera*, *Tabernaemontana corymbosa*, and *Aloe vera* Lupenone, Iboluteine, and (E)-Cycloarta-23-ene-3beta-ol were used in PASS webserver prediction. All the compounds showed the same biological activity. The probability of compounds acting as antineoplastic ranged between 0.823 and 0.955 (Refer to Table 4).

Table 4. Prediction of Biological activities of selected compounds.

Phytocompound	Pa	Pi	Biological activity prediction
Lupenone	0.955	0.004	Antineoplastic
Iboluteine	0.918	0.006	Antineoplastic
(E)-Cycloarta-23-ene-3beta-ol	0.823	0.009	Antineoplastic

Where Pa = probability to be active and Pi = probability to be inactive

4. Conclusions

A total of eighty-one compounds were chosen from three different plants, namely *Tabernaeumontana corymbosa*, *Vitis vinifera*, and *Aloe vera*, for this research. PyRx was used for screening all these phytocompounds, after which ten of them were eliminated owing to a binding affinity of more than -3.5Kcal/mol. The remaining seventy-one compounds were further estimated for their drug likelihood. In total, sixty compounds were found to be drug-likable and underwent testing using a computational approach, i.e., molecular docking studies. For further evaluation, the docked ligands were then compared to the standard inhibitors, i.e., Irinotecan, Curcumin, Doxorubicin, and 5-fluorouracil. Nine phytocompounds possessed higher binding energies than Irinotecan. Among these, the top four compounds were incorporated for bioavailability radar testing. Iboluteine and Ervaluteine of *Tabernaeumontana corymbosa* were observed to be orally bioavailable. Thus, the two compounds of *Tabernaeumontana corymbosa* can be used as a potential plant inhibitor against the targeted

Jab1 protein (graphical abstract). Further *in-vitro* investigations on Iboluteine and Ervaluteine can indeed be fostered in order to identify an efficient and robust treatment for cervical cancer. In all, further *in-vivo* and clinical studies are required to confirm the therapeutic potential of JAB1.

Funding

This research received no external funding.

Acknowledgments

The authors are grateful to the Department of Biotechnology, Delhi Technological University, New Delhi, India, for providing the research facilities and financial support to conduct this study.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. Bhatia, K.; Bhumika.; Das, A. Combinatorial drug therapy in cancer-New insights. *Life Sciences* **2020**, *258*, 118134, https://doi.org/10.1016/j.lfs.2020.118134.
2. Luthra, R.; Datta, S.; Roy, A. Role of Different Peptides for Cancer Immunotherapy. *International Journal of Peptide Research and Therapeutics* **2021**, *27*, 2777-2793, https://link.springer.com/article/10.1007%2Fs10989-021-10289-7.
3. Drețcanu, G.; Iuhas, C. I.; Diaconeasa, Z. The Involvement of Natural Polyphenols in the Chemoprevention of Cervical Cancer. *International Journal of Molecular Sciences* **2021**, *22*, 8812, https://doi.org/10.3390/ijms22168812.
4. Liu, G.; Claret, F. X.; Zhou, F.; Pan, Y. Jab1/COPS5 as a novel biomarker for diagnosis, prognosis, therapy prediction and therapeutic tools for human cancer. *Frontiers in pharmacology* **2018**, *9*, 135, https://doi.org/10.3389/fphar.2018.00135.
5. Pandey, P.; Bajpai, P.; Siddiqui, H.M.; Sayyed, U.; Tiwari, R.; Shekh, R.; Mishra, K.; Kapoor, V.K. Elucidation of the Chemopreventive Role of Stigmasterol Against Jab1 in Gallbladder Carcinoma. *Endocrine, Metabolic & Immune Disorders - Drug Targets* **2019**, *19*, 826-837, https://doi.org/10.2174/1871530319666190206124120.
6. Xiao, H.; Claret, F.X.; Shen, Q. The novel Jab1 inhibitor CSN5i-3 suppresses cell proliferation and induces apoptosis in human breast cancer cells. *Neoplasma* **2019**, *66*, 481-486, https://doi.org/10.4149/neo_2018_181016N772.
7. Torres-Cabala, C.; Li-Ning-Tapia, E.; Hwu, W.-J. Pathology-based Biomarkers Useful for Clinical Decisions in Melanoma. *Archives of Medical Research* **2020**, *51*, 827-838, https://doi.org/10.1016/j.arcmed.2020.09.008.
8. Shackleford, T. J.; Claret, F. X. JAB1/CSN5: a new player in cell cycle control and cancer. *Cell division* **2010**, *5*, 26. https://doi.org/10.1186/1747-1028-5-26.
9. Pandey, P.; Khan, F. Jab1 Inhibition by Methanolic Extract of *Moringa Oleifera* Leaves in Cervical Cancer Cells: A Potent Targeted Therapeutic Approach. *Nutrition and cancer* **2021**, *73*, 2411–2419, https://doi.org/10.1080/01635581.2020.1826989.
10. Pandey, P.; Khan, F.; Maurya, P. Targeting Jab1 using hesperidin (dietary phytocompound) for inducing apoptosis in HeLa cervical cancer cells. *Journal of food biochemistry* **2021**, *45*, e13800, https://doi.org/10.1111/jfbc.13800.
11. Pandey, P.; Khan, F.; Alzahrani, F.A.; Qari, H..; Oves, M. A Novel Approach to Unraveling the Apoptotic Potential of Rutin (Bioflavonoid) via Targeting *Jab1* in Cervical Cancer Cells. *Molecules* **2021**, *26*, 5529, https://doi.org/10.3390/molecules26185529.

12. Park, S.-H.; Kim, M.; Lee, S.; Jung, W.; Kim, B. Therapeutic Potential of Natural Products in Treatment of Cervical Cancer: A Review. *Nutrients* **2021**, *13*, 154, https://doi.org/10.3390/nu13010154.
13. Luthra, R.; Roy, A. Role of Medicinal Plants against Neurodegenerative Diseases. *Current Pharmaceutical Biotechnology* **2022**, *23*, 123-139, https://dx.doi.org/10.2174/1389201022666210211123539.
14. Luthra, R.; Satija, G.; Roy, A. Centellosides: pharmaceutical applications and production enhancement strategies. *Plant Cell, Tissue and Organ Culture (PCTOC)* **2021**, https://doi.org/10.1007/s11240-021-02187-z.
15. Newman, D.J.; Cragg, G.M. Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770–803, https://doi.org/10.1021/acs.jnatprod.9b01285.
16. Luthra, R.; Roy, A.; Pandit, S.; Prasad, R. Biotechnological methods for the production of ginsenosides. *South African Journal of Botany* **2021**, *141*, 25-36, https://doi.org/10.1016/j.sajb.2021.04.026.
17. Roy, A.; Datta, S. Medicinal Plants against Ischemic Stroke. *Current Pharmaceutical Biotechnology* **2021**, *22*, 1302-1314, https://dx.doi.org/10.2174/1389201021999201209222132.
18. Roy, A.; Datta, S.; Bhatia K.S.; Bhumika.; Jha, P.; Prasad, R. Role of plant derived bioactive compounds against cancer. *South African Journal of Botany* **2021**, https://doi.org/10.1016/j.sajb.2021.10.015.
19. Jain, S.; Das, A. Virtual screening of natural compounds as combinatorial agents from Indian medicinal plants against estrogen positive breast cancer. *International Journal of Integrated Education* **2020**, *3*, 266-275, https://doi.org/10.31149/ijie.v3i10.750.
20. Datta, S.; Luthra, R.; Bharadvaja, N. Medicinal Plants for Glioblastoma Treatment. *Anti-cancer Agents in Medicinal Chemistry* **2021,** https://doi.org/10.2174/1871520622666211221144739.
21. Naidoo, C.M.; Naidoo, Y.; Dewir, Y.H.; Murthy, H.N.; El-Hendawy, S.; Al-Suhaibani, N. Major Bioactive Alkaloids and Biological Activities of *Tabernaemontana* Species (Apocynaceae). *Plants (Basel)* **2021**, *10*, 313, https://doi.org/10.3390/plants10020313.
22. Manirakiza, A.; Irakoze, L.; Manirakiza, S. Aloe and its Effects on Cancer: A Narrative Literature Review. *East Afr Health Res J*. **2021**, *5*, 1-16, https://doi.org/10.24248/eahrj.v5i1.645.
23. Abubakar, I.B.; Loh, H.S. A review on ethnobotany, pharmacology and phytochemistry of Tabernaemontana corymbosa. *J Pharm Pharmacol* **2016**, *68*, 423–432, https://doi.org/10.1111/jphp.12523.
24. Salehi, B.; Vlaisavljevic, S.; Adetunji, C.O.; Adetunji, J.B.; Kregiel, D.; Antolak, H.; Pawlikowska, E.; Uprety, Y.; Mileski, K.S.; Devkota, H.P.; Sharifi-Rad, J.; Das, G.; Patra, J.K.; Jugran, A.K.; Segura-Carretero, A.; Contreras, M.M. Plants of the genus Vitis: Phenolic compounds, anti-cancer properties and clinical relevance. *Trends Food Sci Technol* **2019**, *91*, 362–379, https://doi.org/10.1016/j.tifs.2019.07.042.
25. Safaa, F. Phytochemical Constituents from Aloe Vera Leaves and Garden Grass Seeds and Their Utilization as Anticancer Activity. *Int J Eng Res* **2019**, *8*, https://doi.org/10.17577/IJERTV8IS110240.
26. Mohanraj, K.; Karthikeyan, B.S.; Vivek-Ananth, R.P. Bharat Chand, R.P.; Aparna, S.R.; Mangalapandi, P.; Samal, A. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Sci Rep* **2018**, *8*, 4329, https://doi.org/10.1038/s41598-018-22631-z.
27. Brown, N.; Cambruzzi, J.; Cox, P.J.; Davies, M.; Dunbar, J.; Plumbley, D.; Sellwood, M.A.; Sim , A.; Williams-Jones, B.I.; Zwierzyna, M.; Sheppard, D.W. *Big Data in Drug Discovery* **2018**, *57*, 277–356, https://dx.doi.org/10.1016/bs.pmch.2017.12.003.
28. Dallakyan, S.; Olson, A.J. Small-Molecule Library Screening by Docking with PyRx. **2015,** 243–250, https://doi.org/10.1007/978-1-4939-2269-7_19.
29. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Scientific Reports* **2017**, *7*, 42717, https://doi.org/10.1038/srep42717.
30. Lipinski, C.A. Lead- and drug-like compounds: the rule-of-five revolution. *Drug Discov Today Technol* **2004**, *1*, 337–341, https://doi.org/10.1016/j.ddtec.2004.11.007.
31. Morris, G.M.; Huey, R.; Lindstrom, W.; Sanner, M.F.; Belew, R.K.; Goodsell, D.S.; Olson, A.J. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *Journal of Computational Chemistry* **2009**, *30*, 2785-2791 https://doi.org/10.1002/jcc.21256.
32. Maurya, P.; Pandey, P.; Khan, F.; Mishra, R.; Chaudhary, R.; Singh, S. K. Study to Elucidate the Inhibitory Potential of Selected Flavonoids against Jab1 in Cervical Cancer. *Biointerface Research in Applied Chemistry* **2021,** *12*, 1290-1303**,** http://dx.doi.org/10.33263/BRIAC121.12901303.

Exploring intergenerational differences in the virtue of appreciation at the workplace

Naval Garg and Manju Mahipalan

Abstract

Purpose – *The present study aims to explore the intergenerational differences in the virtue of appreciation. The construct of appreciation takes a broadened perspective with six dimensions of "have" focus, awe, ritual, present moment, loss/adversity and interpersonal appreciation. Four different generations are studied – baby boomers (above 55 years), pre-millennial or Generation X (41–55 years), millennial or Generation Y (23–39 years) and post-millennial or Generation Z (below 23 years).*

Design/methodology/approach – *The collected data is analyzed in two stages. First, the adaptability of the scale is examined using exploratory factor analysis, confirmatory factor analysis, reliability and validity estimates. And in the second stage, the variations in appreciation scores are explored using ANOVA and post hoc analysis.*

Findings – *The results reveal statistically significant intergenerational differences among four subscales of appreciation, i.e. "have" focus, awe, present moment and loss/adversity. Only one subscale of appreciation, i.e. ritual, does not vary significantly across respondents of different generations. Also, younger generations have lesser scores on subscales of appreciation than older generations. Thus, it seems that older generations are generally more appreciative than younger generations.*

Originality/value – *The present study, to the best of the authors' knowledge, could be the first research that examines the variation among generations for the experience of broader construct of appreciation within the Indian context.*

Keywords *Appreciation, Baby boomers, Millennial, Generational differences*

Paper type *Research paper*

Naval Garg is based at the University School of Management and Entrepreneurship, Delhi Technological University, New Delhi, India.
Manju Mahipalan is based at the Department of Management Studies, NIT Tiruchirappalli, Tiruchirappalli, India.

Introduction

Workforce diversity is all-pervasive in the 21st century. Every organizations – public or private, manufacturing or service, profit or non-profit orientated firms, Indian or multi-national companies (MNCs) – witness the celebrated presence of a diverse workforce. It refers to hiring and promoting employees of a different race, gender, culture, age, religion, ethnicity, education, caste, experience, interest, personality, cognitive style, economic status and functional diversity (Manoharan and Singal, 2017; Li *et al.*, 2020; Choi and Rainey, 2010; Tamunomiebi and John-eke, 2020). A diverse workforce is desirable from both social justice and business perspectives (Tamunomiebi and John-eke, 2020). On the social front, it facilitates equal and fair opportunities for all and eliminates stereotyping and prejudice against a specific group or category of people. Workforce diversity is also recommended as an indirect persuasion towards an egalitarian society (Garg *et al.*, 2021). From a business perspective, employee diversity promotes diverse ideas, perspectives, ways of working and cognition styles, which breed innovation and creativity. It also acts as a deterrent to groupthink and group polarization processes. Workforce diversity leads to job satisfaction (Choi and Rainey, 2013), job performance (Vanderschuere and Birdsall, 2018), better skill utilization (Pink-Harper and Rauhaus, 2017), positive work attitude and behaviour (Ju and Li, 2019) and job–person fit (Li *et al.*, 2020).

Received 17 July 2021
Revised 13 December 2021
15 March 2022
Accepted 2 May 2022

© Emerald Publishing Limited, ISSN 1747-1117

Among various other bases of diversity, modern organizations experience a blend of employees from different generational cohorts. The generational cohorts are differentiated based on their year of birth. It is suggested that people of one generation experience similar life situations, values, events, technology and social changes (Rank and Contreras, 2021). These shared experiences influence their attitude, outlook and behaviours in both personal and professional life. Researchers have concluded that different generations develop distinct differences and characteristics (Rank and Contreras, 2021; Twenge, 2010; Rudolph *et al.*, 2020; Lang, 2020). Previous studies have reported significant generation-based variations in employee's attitudes (Alferjany and Alias, 2020), homesickness (Garg *et al.*, 2021), motivational drives (Wong *et al.*, 2008), learning orientation (D'Amato and Herzfeldt, 2008), work motivation (Deal Stawiski *et al.*, 2013), career types and success (Dries *et al.*, 2008), etc. Although adequate studies have explored generational differences at the workplace, there is a call to examine significantly internalized, personalized and latent traits and emotions such as spirituality, gratitude, appreciation, compassion and happiness among employees of different generational cohorts.

The current study is a humble attempt to understand how appreciation plays out for different generations at the workplace. Appreciation is the quality of noticing and recognizing the goodness in people and things around us. Being appreciative cheers our spirits up and results in a perceptible improvement in awareness so that we live a more meaningful and mindful life. People who are not appreciative might only focus on what they lack and remain bitter and resentful. As with the case of other emotions, appreciation can be conceptualized as an emotion or a disposition. While appreciation as an emotion is a momentary response to an event and is brief, dispositional appreciation enables a person to feel appreciation quite often and in various circumstances. Adler and Fagley (2005), who have carried out seminal works on the construct of appreciation, defined appreciation as "acknowledging the value and meaning of something—an event, a person, a behaviour, an object—and feeling a positive emotional connection to it" (p. 81). They further argued that appreciation consists of several cognitive and emotional elements which seem important in understanding an individual's experience of life satisfaction and happiness. Additionally, appreciation facilitates the formation and maintenance of social bonds (Adler and Fagley, 2005; Algoe, 2012; Algoe *et al.*, 2008; Kubacka *et al.*, 2011). Appreciation has also been studied in relation to well-being by creating positive affect and mental health (Steindl-Rast, 2004; Fagley, 2018; Lim, 2017).

A motivation to undertake the study is the possibility that there are differences in the ways people experience and express appreciation, which could subsequently affect the quality of their personal and professional life spaces. Joshi *et al.* (2010) assumed that multiple generations co-exist at workplaces and interact with each other to accomplish common organizational goals. Rudolph and Zacher (2018) noted that the organizational workforce is becoming increasingly age-diverse and contemplated that age-related differences can influence individual-level behaviours. The social constructionist perspective of generation moves a step ahead, urging the researchers to manage perceptions of generations and their differences at workplaces (Rudolph *et al.*, 2020). Cogin (2012) pointed out that little attention has been paid to age diversity in the workplace compared to gender and cultural diversity. They also provided evidence from multiple countries to establish the legitimacy of intergenerational differences as an essential social categorization variable. Although many researchers have suggested significant intergenerational variations, differences in the virtue of appreciation have hardly been studied previously. Given the importance of appreciation in an individual's personal and professional well-being, it is imperative to investigate the level of employees' appreciativeness. In this frame of reference, the present study presents itself with the following objective: to explore the intergenerational differences in the virtue of appreciation among employees. Four generations of Indian employees, i.e. babyboomers, pre-millennial, millennial and post-millennial, are studied.

Also, previous researchers have opined that appreciation is a cultural-specific virtue. It has a common core with culturally ubiquitous features (Morgan *et al.*, 2014). Also, the experience, intensity, direction and expression of gratitude may vary with culture and nationality (Garg and Mehak, 2021; Sommers and Kosmitzki, 1988; Tudge *et al.*, 2015). The present study explores the virtue of appreciation in the Indian context. Indian cultural ethos presents one of the most comprehensive and elaborated conceptualizations of appreciation. It seems to be premised in the Indian notion of Rna or pious obligations or sacred duties. These duties are considered essential for an individual's spiritual upliftment and growth (Pandey and Navare, 2018). These pious obligations are Manuṣya-yajña (towards fellow human beings), Deva-yajña (towards God and nature), Pitr-Yajna (towards the family, ancestors and cultural values), Brahma-yajña (towards knowledge, skills and talents) and Bhūta-yajña (towards the eco-system). These *rnas* encourage people to remain appreciative of ancestors, friends, nature, God and skills for their valuable contributions in one's life. In a nutshell, it aspires one to remain appreciative for everything, everywhere and every time.

The proposed exploration of intergenerational differences will reveal whether younger generations are more appreciative than their previous generations. It is important to note that with the advancement of science and technology, comfort has risen exponentially. Today's generation has almost everything to save themselves from the wraths of nature. It would be interesting to see differences in appreciation in light of growing economic prosperity and growth.

The present paper is broadly divided into four parts. First, the concept of appreciation and its various constructs are discussed. Second, intergenerational differences are explained with the help of available literature. Third, the adaptability of the appreciation scale is examined using exploratory factor analysis, confirmatory factor analysis (CFA), reliability and validity estimates. Also, the variations in appreciation scores are explored using ANOVA and post hoc analysis. And last, the findings of the study are discussed in light of previous studies along with theoretical and practical implications of the study.

Literature review

Appreciation and its constructs

There have been debates on defining the construct of appreciation, and it has been used synonymously with the term gratitude (Lambert *et al.*, 2009b; Wood *et al.*, 2007). The majority of the studies have focused on the construct of gratitude as an emotion felt towards a benefactor as a response to the benefit or assistance received, i.e. state-like gratitude or "benefit-triggered" gratitude (Emmons and McCullough, 2003). The problem of conceptual clarity arises when certain expressions of gratitude may not necessarily involve a benefactor. These expressions represent a "higher order" of gratitude that spans many aspects of life. A handful of studies provide cogent arguments for the existence of distinct forms of gratitude. For instance, a prototype analysis of gratitude showed that people have varied notions of gratitude, which are not addressed in the current definitions of gratitude (Lambert *et al.*, 2009a). Sansone and Sansone (2010) restated that gratitude could have multiple interpretations in various contexts and defined it as a general state of appreciation. Wood *et al.* (2010) presented a life orientation conception of gratitude, implying that there are different ways of viewing and experiencing the world for a grateful personality. Targeted gratitude or triadic gratitude is about one person being grateful to the other. And generalized or propositional or dyadic gratitude involves gratefulness for the state of affairs (Manela, 2016; McConnell, 2020). Manela (2019) further noted that appreciation is a response to the beneficial or valuable states of affairs constituted by cognitive elements. Researchers have confirmed these discrepancies in definitions and underscored the necessity for accommodating all forms of gratitude (Hlava and Elfers, 2014; Steindl-Rast, 2004). Thus,

appreciation is regarded as a broader construct encompassing all forms of gratitude and several other dimensions.

Adler and Fagley (2005) conflated eight different aspects to elucidate the concept. These are "have" focus, awe, ritual, present moment, self/social comparison, gratitude, loss/adversity and interpersonal appreciation. The "have" focus implies focusing and valuing what one possesses instead of playing them down. It can also mean a lack of sense of deprivation or scarcity. People with a sense of abundance can continually focus on gratitude and expand it. The awe aspect is the feeling of wonder, admiration or reverence for the beauty one sees in life or nature. Keltner and Haidt (2003) offered a prototype perspective of awe. They define it as a complex emotion arising from a perception of vastness and need to accommodate the perception mentally. Chirico and Yaden (2018) discussed the transcendental and transformative aspects of awe. They suggested that it might restructure the individual's inner world at the most intimate level. Ritual refers to rituals or practices one indulges in to foster the feeling of appreciation.

Hobson *et al.* (2017) considered ritual as a unique and remarkable aspect of human behaviour. And they proposed that rituals could alleviate emotional distress among people, regulate performance and social connections. The present moment is engaging or living at the moment with complete awareness and taking note of the silver lining in every situation. In this context, considerable attention has been paid to mindfulness, which involves a receptive, non-judgmental awareness of and attention to whatever is present in the moment (Kiken *et al.*, 2017). Self/social comparison refers to the comparison of the present moment or circumstance either with past moments in own life or with others to reinforce the feeling of appreciation and remind oneself to be appreciative about life irrespective of the situation. Self/social comparisons can influence a person's affective thinking. Nicuă and Constantin (2021) proposed that downward comparative thinking can elevate gratitude when people appreciate their situation more once they compare themselves with others who are in less advantageous positions.

Gratitude is the gratefulness directed towards others for the help or benefits received. The aspect of gratitude discussed here is benefit-triggered gratitude. Rusk *et al.* (2015) explained this as an individual's appraisal of benefit that concerns the perceived agency of another person or entity. Loss/adversity signifies the experiences of loss or adversities, which may nurture an appreciation of what one still possesses through a change in perspective. Popovich's (2014) study demonstrated that losses strengthened people's sense of gratitude, and many viewed their lives as a precious gift when losses served as a reminder to be grateful. Interpersonal appreciation is noticing and valuing others' presence in life and expressing the same. This aspect of appreciation spans beyond benefit-triggered gratitude, where people are grateful to others' presence in their lives as a whole. When people have this form of appreciation for others, they can tune in to the thoughts, emotions and motives of others' positive actions and focus on understanding and mirroring other people's kindness (Bono and Sender, 2018).

Adler and Fagley (2005) demonstrated that strong correlations existed among various aspects of appreciation and positive affect and life satisfaction. Positive associations were also found between appreciation, happiness and sensitivity to reward (Tucker, 2007). Watkins *et al.* (2003) operationalized gratitude in terms of an individual's sense of abundance, interpersonal appreciation, ability to recognize and appreciate simple pleasures in life. These dimensions of gratitude overlap with various aspects of the current conceptualization of appreciation. The gratitude thus operationalized was found to be significantly related to subjective well-being. Watkins (2004) deliberated the plausible relationship between gratitude and happiness. He also suggested that gratitude enables individuals to focus more on the blessings (the have focus), recollect the pleasant life events (self-comparison), find benefits even in unpleasant circumstances (loss/adversity) and enhance the quality of relationships (interpersonal appreciation). Following the suit,

Lin (2015) posited that gratitude does not have a unifactorial structure and proposed a higher order construct encompassing some of the above-mentioned aspects of appreciation, such as have focus, present moment and loss/adversity. He found these constructs associated with higher subjective well-being, psychological well-being, self-esteem and lessened depression.

Besides, literature furnishes limited yet promising evidence for the various aspects of appreciation discussed here. For instance, Simons *et al.* (2019) demonstrated that a sense of abundance was associated with momentary positive and negative affect, thereby revealing an affective pathway between appreciation and well-being. Piff *et al.* (2015), in an experimental study, found that awe resulted in greater generosity and prosocial values. Awe also led to increased ethical decision-making while reducing the sense of entitlement. Stellar *et al.* (2018) proposed that awe is a self-transcending emotion that fosters connectivity among people and promotes prosocial behaviour. Likewise, an experimental study found that enacting rituals reduced the anxiety levels in individuals by calming them down and improving their overall performance (Brooks *et al.*, 2016). Hobson *et al.* (2017) presented that rituals can create feelings of transcendence. It also allows a person to move beyond ego-based thoughts and anxieties. Paying attention to the present moment was found to be positively related to an increased level of life satisfaction through abating negative rumination (Felsman *et al.*, 2017). A similar concept of mindfulness also has been related to self-regulation of thoughts, emotions and behaviour, which is linked to performance and well-being (Glomb *et al.*, 2011). Gratitude has been studied extensively and linked to many positive outcomes in individuals. Algoe *et al.* (2008) evinced the social functions of gratitude such as relationship formation, which is more than a feeling of gratefulness to a benefactor or merely repaying the benefits. Benefit-triggered gratitude is linked to prosocial behaviours (Fehr *et al.*, 2017; Ma *et al.*, 2017), maintenance behaviours (Kubacka *et al.*, 2011) and overall well-being of an individual (Watkins *et al.*, 2003; Wood *et al.*, 2010).

Intergenerational differences among employees A generational group or cohort is an "identifiable group that shares birth years, age location and significant life events at critical development stages" (Kupperschmidt, 2000, p. 66). Macky *et al.* (2008) suggest that differences between generations happen majorly because of environmental influences on the development of personality, values, beliefs and expectations. There are dissenting views on the actualities of generational differences. An early study attempting to divulge generational differences in individuals' personalities and motivational drives produced no affirmative results (Wong, *et al.*, 2008). Parry and Urwin (2011) observed that the empirical evidence for generational differences is mixed, and there are no telling results that present generation as a significant factor in the organizational context. Lester *et al.* (2012) highlighted the stereotyping of generational differences by demonstrating that the perceived generational differences outnumbered actual differences. Lyons and Kuron (2014) addressed the subject with a similar sentiment, stating that studies in the domain are primarily descriptive, methodologies are inconsistent and evidence paradoxical. A meta-analysis review could not render substantial evidence to support meaningful differences among generations on work-related attitudes (Costanza *et al.*, 2012). Likewise, Becton *et al.* (2014) noted no major generational differences in workplace behaviour instead of the popular stereotypes. A recent study concluded that there seemed to be more similarities than differences in generations for workplace attitudes as the differences reported were paltry (Cucina *et al.*, 2018). On the flip side, a handful of studies have emphasized and explored the bearing of generational differences on various aspects of work–life. For instance, Gursoy *et al.* (2008) found significant generational differences in world views and perspectives on work among the hospitality workforce.

Twenge's (2010) review of the empirical evidence on generational differences found variations among generational cohorts concerning intrinsic and extrinsic work values, work

ethic and centrality and perceptions about leisure. Similar results were obtained for nurses where distinct generations prioritized personal and professional values differently (Şenyuva, 2018). Schullery (2013) showed that generational differences are noticeable in the case of values at work, and when conflated, they had a considerable impact on the engagement levels. Similar results were brought forth when generations differed in the vigor, dedication and absorption aspects of job engagement (AkhavanSarraf *et al.*, 2017). Campbell *et al.* (2015) argued that generational grouping is a valuable tool for making meaningful distinctions among people as well as predicting their behaviour. It was also found that generational cohorts differ in their meaning-making process about work and listing the sources of meaningful work (Weeks and Schaffert, 2019). Several longitudinal, cross-sectional, meta-analytic studies are adduced to prove the concept validity of generation as a workplace variable and nuanced and methodologically rigorous understanding of the phenomenon is called for (Lyons and Kuron, 2014).

Joshi *et al.* (2011) identified multiple perspectives on generations in organizations which include generations based on chronology and genealogy. A chronological approach based on age relies on the assumption that a set of attitudes, values and preferences shared by individuals similar in age is different from individuals belonging to another age cohort. Age as a critical differentiation criterion is an oversimplification of a complex concept. Nevertheless, a considerable amount of research has worked with this limitation (Andert, 2011; Kowske *et al.*, 2010). Generational differences could be country-specific; however, Indian culture is not unitary or homogeneous as it is highly diverse, which is attributable to various socio-economic and linguistic factors (Srinivasan, 2012). Some authors have attempted to define generations at the Indian workplace using a global framework (Chaudhuri *et al.*, 2019; Erickson, 2009; Ghosh and Chaudhuri, 2009; Srinivasan *et al.*, 2014; Srinivasan *et al.*, 2011). Thus, because of the heterogeneous nature of Indian culture as well as the lack of any established cohorts, the current study also makes use of the age-based construct of generation and includes four age cohorts, namely, baby boomers, Generation X, Generation Y and Generation Z, based on birth years(Weeks and Schaffert, 2019). Baby boomers were born between 1946 and 1964, Generation X between 1965 and 1983, Generation Y or Millennials between 1984 and 2002 and Generation Z after 2002. Few studies have explored the intergenerational differences for the construct of appreciation. Hitherto only one article has studied the age differences and experience of gratitude from a developmental perspective (Chopik *et al.*, 2019). Probably, no other study has examined the intergenerational differences in appreciation and its construct among professional employees.

Methodological framework

Objective and sample

The present study explores the intergenerational differences in the virtue of appreciation. Four different generations are studied: baby boomers (above 55 years), pre-millennial or Generation X (41 –55 years), millennial or Generation Y (23–39 years) and post-millennial or Generation Z (below 23 years). Six subscales of appreciation, namely, “have” focus, awe, ritual, present moment, loss/adversity and interpersonal, are studied along with overall appreciation level. The study is based on a sample of 630 employees engaged in banks, insurance companies, IT firms and manufacturing companies. The respondents were randomly contacted near the metro station, bus stand and railway station in Delhi-National Capital Region. A total of 685 respondents were approached. Around 55 responses were discarded for being incomplete. The sample comprises 310 males and 320 females. Among participants, 87 are baby boomers, 193 are pre-millennial, 240 are millennial and the remaining 110 are post-millennial. Also, respondents belonging to categories with yearly income “below 3 lakhs”, “3–5 lakhs”, “5–10 lakhs”, “10–15 lakhs” and “above 15 lakhs” categories are 176, 122, 121, 104 and 107, respectively.

Measures

Appreciation scale Adler and Fagley's (2005) 18-item Appreciation Scale (short form) is used to measure the respondents' appreciation level. The scale accesses six dimensions of appreciation, namely, "have" focus (five statements), awe (three statements), ritual (three statements), present moment (four statements), loss/adversity (two statements) and interpersonal (one statement). These statements are rated on a seven-point rating scale ranging from 1 (never), 2 (a few times in my life), 3 (about once a year), 4 (about once in a month), 5 (about once a week), 6 (about once a day) and 7 (more than once a day). The items include, "I count my blessings for what I have in this world", "I remind myself how fortunate I am to have the privileges" (for "have" focus), "I have moments when I realize how fortunate I am to be alive", "I reflect on how lucky I am to be alive" (for awe), "I give thanks for something at least once a day", "I do things to remind myself to be thankful" (for ritual), "I recognize and acknowledge the positive value and meaning of events in my life", "I remind myself to appreciate the things around me" (for the present moment), "The problems and challenges I face in my life help me to value the positive aspects of my life", "Thinking about dying reminds me to live every day to the fullest" (for loss) and "I remind myself to appreciate my family" (for interpersonal).

Data analysis and results

The collected data is analyzed in two stages. First, because the appreciation scale has been developed and validated in the Western contexts, the adaptability of the scale is examined using exploratory factor analysis, CFA, reliability and validity estimates. Structure equation modelling (SEM) using CFA is preferred over partial least squares structural equation modeling (PLS-SEM) for two reasons. First, PLS-SEM is generally used to explore the model (Fan *et al.*, 2016). It is preferred when the theoretical underpinning of the model is not clear. However, the present study required confirmation of the already theoretically premised model. First, Bentler and Chou (1987) suggested two prerequisites for the application of CFA: the prior knowledge of indicators for a latent variable, and the previous understanding of the relationships between the latent variable and its indicators. Because both conditions are met here, CFA is used in this study, And in the second stage, the variations in appreciation scores are explored using ANOVA and post hoc analysis.

Exploratory and confirmatory factor analysis

The participants are divided into two sub-samples. Exploratory factor analysis using varimax rotation is carried out using the first sample of 315 participants in SPSS. Table 1 concludes satisfactory factor loading (greater than 0.60) of all statements except Item 18. This item is discarded for further analysis. The loading of the remaining 17 items ranges from 0.604 to 0.825. The items of appreciation could also explain a cumulative variance of 63.016. After deriving acceptable factor loading for all statements, the second sample is subjected to model fit analysis through CFA in AMOS. Standards model fit indices are root mean square error of approximation (RMSEA) < 0.08, goodness of fit indices (GFI) ≥ 0.90, comparative fit indices (CFI) ≥ 0.90, normative fit indices (NFI) ≥ 0.90 and standardized root mean square residual (SRMSR) < 0.08 (Bentler, 1990; Kline, 2009).

Default model provides reasonably fit indices, RMSEA = 0.04, GFI= 0.94, CFI = 0.97, CFI= 0.91 and SRMSR = 0.081. All values except SRMSR indicate a good model fit. Modification indices are examined to improve the model. Accordingly, two statements of "haves" focus ("I really notice and acknowledge the good things I get in life" and "remind myself how fortunate I am to have the privileges and opportunities I have encountered in life") allowed to correlate. Thereafter, results conclude a better model fit, RMSEA = 0.04, GFI= 0.95, CFI = 0.98, CFI= 0.94 and SRMSR = 0.052. Now, after EFA and CFA, the reliability and validity of the scale are examined.

Table 1 Factor structure of appreciation scale

Items	*Loading*	h^2
I count my blessings for what I have in this world	0.753	0.621
I remind myself how fortunate I am to have the privileges and opportunities I have encountered in life	0.821	0.762
I reflect on how fortunate I am to have basic things in life like food, clothing, and shelter	0.795	0.708
I really notice and acknowledge the good things I get in life	0.825	0.797
I remind myself to think about the good things I have in my life	0.806	0.712
I have moments when I realize how fortunate I am to be alive	0.751	0.651
I reflect on how lucky I am to be alive	0.812	0.702
I feel that it is a miracle to be alive	0.740	0.616
I give thanks for something at least once a day	0.812	0.729
I do things to remind myself to be thankful	0.793	0.695
I believe it is important to remind myself to be thankful for things on a consistent basis	0.631	0.596
I enjoy the little things around me like the trees, the wind, animals, sounds, light, etc.	0.630	0.621
I recognize and acknowledge the positive value and meaning of events in my life	0.675	0.638
I remind myself to appreciate the things around me	0.609	0.650
When I stop and notice the things around me, I feel good and content	0.604	0.628
The problems and challenges I face in my life help me to value the positive aspects of my life	0.665	0.523
Thinking about dying reminds me to live every day to the fullest	0.615	0.521
I remind myself to appreciate my family	0.345	0.440
Number of items		18
Variance explained (%)		63.016

Source: Primary data, h^2: communalities

Reliability and validity

The reliability of the scale is studied with the help of Cronbach's alpha (CA) and composite reliability (CR) values. Hair *et al*. (1998) claimed that a value of 0.70 and above is required to establish good reliability of the scale. Table 2 reports that CA and CR values for "have" focus, awe, ritual, present moment and loss/adversity are 4.76, 4.57, 5.15, 5.29 and 5.33, respectively. These values are appropriate to conclude the reliability of the appreciation scale for Indian samples. Average variance explained (AVE) values are used to check the scale's convergent and discriminant validity. Fornell and Larcker (1981) recorded that a value of 0.50 and above concludes convergent validity. The AVE values highlighted in Table 3 provide empirical evidence of the convergent validity of the scale. The Fornell–Larcker discriminant validity testing system demands a higher value of the square root of AVE values than correlation coefficients to establish discriminant validity. The results highlighted in Table 3 conclude the discriminant validity of the scale. Table 3 also depicts the correlation matrix of five subscales of appreciation. Although the distinctiveness of these subscales has been ensured using CFA, the issue of multicollinearity is addressed with the help of variance inflation index (VIF) values. The VIF values are 2.232, 2.116, 1.794, 2.122 and 2.192 for have focus, awe, ritual, present moment and loss/adversity. These values negate any possibility of multicollinearity among subscales of the appreciation scale.

Table 2 Descriptive statistics, reliability and validity estimates

Variable	*Mean*	*SD*	*CA*	*CR*	*AVE*
"Have" focus	4.76	1.55	0.921	0.854	0.640
Awe	4.57	1.49	0.825	0.811	0.590
Ritual	5.15	1.44	0.770	0.791	0.562
Present moment	5.29	1.24	0.811	0.733	0.587
Loss	5.33	1.35	0.786	0.785	0.579
Overall	5.02	1. 34	0.837	0.950	0.533

Notes: SD: standard deviation; CA: Cronbach's alpha; CR: composite reliability; AVE: average variance explained
Source: Primary data

Table 3 Correlation matrix and discriminant analysis results

Variable	*1*	*2*	*3*	*4*	*5*
"Have" focus	*1(0.08)*				
Awe	0.663**	*1(0.768)*			
Ritual	0.583**	0.562**	*1(0.749)*		
Present moment	0.474**	0.457**	0.478**	*1(0.766)*	
Loss	0.361**	0.336**	0.427**	0.651**	*1(0.760)*

Notes: **Significant at 0.05; square root of AVE values indicated in diagonal (in italics)
Source: Primary data

Intergenerational variations

The intergenerational variations in overall appreciation and its five subscales are examined using ANOVA. Table 4 concludes significant variations in overall appreciation scores among respondents of different generations. It is interesting to note that younger generations are relatively less appreciative than older generations (appreciation score, post-millennial = 5.22, millennial = 5.13, pre-millennial = 5.03 and baby boomers = 5.22). It also reveals statistically significant intergenerational differences among four subscales of appreciation, i.e. "have" focus, awe, present moment and loss/adversity. Only one subscale of appreciation, i.e. ritual, does not vary significantly across employees of different generations. Also, younger generations have lesser scores on subscales of appreciation than older generations. Thus, it seems that employees of older generations are generally more appreciative than that of younger generations.

Further, Tukey's honestly significant difference post hoc analysis is used to have a better picture of generation-based differences in the overall appreciation scale. First of all, homogeneity of variance is explored to satisfy basic assumptions. Then, Tukey's test is

Table 4 Generation-based variations in appreciation score

Appreciation subscale	*Generation*	*Mean*	F-*value*	*Sig.*
"Have" focus	Post-millennial	4.61	4.521	0.012**
	Millennial	4.71		
	Pre-millennial	4.85		
	Baby boomer	5.07		
Awe	Post-millennial	4.37	2.823	0.038**
	Millennial	4.53		
	Pre-millennial	4.78		
	Baby boomers	4.82		
Ritual	Post-millennial	5.03	1.143	0.343
	Millennial	5.12		
	Pre-millennial	5.28		
	Baby boomer	5.31		
Present moment	Post-millennial	5.07	3.208	0.023**
	Millennial	5.35		
	Pre-millennial	5.38		
	Baby boomer	5.49		
Loss/adversity	Post-millennial	5.13	3.124	0.027**
	Millennial	5.31		
	Pre-millennial	5.43		
	Baby boomer	5.47		
Overall appreciation	Post-millennial	4.84	5.270	0.005**
	Millennial	5.03		
	Pre-millennial	5.13		
	Baby boomer	5.22		

Note: **Significant at 0.05
Source: Primary data

applied using SPSS. A statistically significant variation is observed between baby boomers and the other three generations, i.e. pre-millennial, millennial and post-millennial. Similarly, pre-millennial also report significant differences from the other three generations. Only the variation between millennial and post-millennial is reported as statistically insignificant.

Discussion

Appreciation has expansive scope both at micro and macro levels because, as a positive emotion, when nurtured and regulated, appreciation can result in perdurable benefits for individuals as well as organizations. To gain these benefits, the focus should be on bringing into play the kind of organizational culture built on the whole gamut of positive emotions, including appreciation. Researchers have touched on ways in which appreciation can be fostered at workplaces. These include primarily appreciation-centered HR initiatives, various interdependent work practices and gratitude interventions (Fehr *et al.*, 2017; Kaplan *et al.*, 2014). How to leverage these interventions or practices to foster a culture based on the virtue of appreciation should be discussed against the backdrop of the intergenerational differences evident in this study. Today's organizations witness multiple generations working side-by-side with differing perceptions, values and work styles. Moreover, this composition is evolving further, with more individuals from the younger generations joining the workforce.

The "peer personality" is a sum of collective attitudes about family, gender roles, institutions, religion, lifestyles and visions about the future. For the same reasons, personalities from distinct generations may connect or confront, attract or repel each other (Strauss and Howe, 1991). This personality is moulded by historical, cultural, social and life experiences, which influence the individual's emotions and values by which they function in the society, including workplaces (Zopiatis *et al.*, 2012). In short, the effectuality of these practices is contingent upon numerous contextual, situational and personal factors. People vary in their levels of experiencing and expressing appreciation. For instance, except for one aspect of ritual, intergenerational differences are evident for all subscales of appreciation. People belonging to the older generation were more appreciative than people in younger generations. The latest study has demonstrated affirmative evidence of the impact of generational diversity on spirituality and religion at the workplace. McMurray and Simmers (2020) found that baby boomers and Generation X displayed higher levels of spirituality than millennial. These results seems in line with previous studies conducted by various researchers, including Twenge (2010), Schullery (2013), Campbell *et al.* (2015) and AkhavanSarraf *et al.* (2017). These scholars have suggested significant generational differences in work values, work ethic, personal and professional values, vigor, dedication and absorption aspects of job engagement among employees of different generations. The findings also might sparsely validate the Campbell *et al.* (2015) generational difference model of extrinsic individualism, which argues that generations, as they evolve, will have more extrinsic self-focus and less inward focus, among other things.

Practical implications of the findings

The present study's finding corroborates the influence of generational diversity on appreciation, which is reciprocally related to the construct of spirituality (Fagley and Adler, 2012). Times have changed, and worldviews have shifted. Organizations need to reckon with this fact and rely on felicitous strategies that benefit people of all generations. For instance, because the younger generation of millennials and post-millennial do not appreciate the relevance of the present moment or what they possess to the extent to which older generations do, training programs or interventions designed should assist people to be more mindful and aware about the present and encourage them to be grateful for what they have and rise above discouragement, doubt and envy. Assuming that all people begin

from the same point and rolling out an undetailed blanket program for everyone might turn counterproductive.

The findings of the present study recommend that an organization may try to institutionalize appreciation-based organizational culture. HR managers may try to initiate practices such as appreciation week, appreciation newsletters, appreciation meetings (just like quality circles), seminars and workshops on appreciation, to facilitate an appreciative workforce. Also, appreciation, like other positive emotions, is contagious, and an act of appreciation by top management and leaders is highly infectious. Thus, top management or managers may start practicing appreciating organizational vision and mission, working culture and artifacts of the firms, subordinates and colleagues. It may have a cascading effect, and employees in lower hierarchical rungs may automatically develop the virtue of appreciation. Besides, older employees may act as a mentor for younger employees. It may help in the development of a sense of appreciation among younger ones. With the help of an organizational development expert, an organization may experiment with specific interventions such as a buddy system, multigenerational teams, periodic T-training, experience sharing sessions of older employees, etc.

The findings are significant for global MNCs working in India. Employment engagement and loyalty remain significant concern for any company trying to establish its business in a foreign country. Employers of these MNCs may use various techniques to increase the appreciation levels of the employees. The findings suggest that employer requires different interventions to elicit appreciation among employees of different generations. The Indian notion of *rnas* may be rekindled among younger generations for the desired results.

Theoretical contributions

The present study contributes to the existing literature in three ways. Firstly, it significantly adds to the fractured stream of evidence that supports the factuality of intergenerational differences at the workplace. The study and its findings are a response to the implication of having additional research to ascertain the existence of generational differences proposed by many researchers. The results hold up the assumption that generational groupings do exist, and they are useful in explaining perceptual, attitudinal and behavioural differences among people. Secondly, this research has centered on appreciation, a construct that is not sufficiently mapped in the domain of management. As stated in the earlier sections of this article, appreciation is an inclusive concept of gratitude and a key factor for workplace success. Gratitude takes a broader definition of appreciation with eight diverse aspects. A higher order conceptualization of gratitude has been proposed and empirically tested before, but further corroboration of the same is always warranted, especially when research on positive emotions is incipient. Researchers from different cultural background may also try to validate this higher order model of gratitude. Thirdly, only a few studies inspect the generational differences in the experience of positive emotions and, in particular, appreciation. The findings are indicative of the fact that generational cohorts have their own unique set of values that are displayed at work. They vary in their attitudes, behaviour and preferences at work, and they need to be managed differently. The findings also might sparsely validate Campbell *et al.*'s (2017) generational difference model of extrinsic individualism in Indian settings. The validation of a model in different cultures is essential for universalization of a theory. The findings of the study will help in global acceptance of the model. In summary, the organization should strive to build a culture that would foster the virtue of appreciation. Understanding generational differences and responding to them effectively will allow different generations to increase their appreciation of one another and minimize conflicts.

Limitations, scope for future studies and conclusion Although the study renders some interesting insights on the virtue of appreciation which has been inadequately examined, it is not without limitations. The study is cross-sectional in design which may be considered as

a disbenefit to the exploration of a construct such as appreciation. However, no temporal links are investigated, and the objective was to arrive at a contemporary snapshot of the differences in the experience of appreciation among generations. Secondly, the generalizability of results is constricted to some extent. There are differences each generational group might have, besides age known for cohort effects, which are beyond the scope of this study. Finally, there could be problems which are associated with self-reported data which may not capture complex social process in totality. Survey data is also susceptible to response sets where participants record their responses in a biased fashion. The results presented have only scratched the surface of a research area that is vast and contains multitudinous possibilities. To begin with, researchers can further probe and conceptualize the construct of appreciation by building on contextual, situational and, more importantly, cultural dimensions. The nexus between intergenerational differences in appreciation and other workplaces attitudinal and behavioural variables using longitudinal models could be another potential research direction. Apart from age, personality differences in the experience of appreciation could be studied in detail, which would provide more impactful insights into human behaviour. Examining appreciation as an intervention mechanism in different organizational settings could underpin some of the theoretical propositions put forward by researchers so far.

Research in the domain of positive psychology is on the rise primarily because it enlarges our understanding of optimal human functioning. The present research is a progressive and significant addition to this burgeoning area of interest. Appreciation, as conceptualized in this study, captures a wide spectrum of emotions associated with gratefulness. Generational differences in the experience of this virtue bring up the differing worldviews of people, the implications of which are consequential. As far as organizations are concerned, this is crucial to cope with the dynamics of changing social relations. Organizations need to keep intact the divergent workforce and minimize intergenerational conflicts if any. As proposed earlier, building on the resource of appreciation is one of the practical and realistic ways in which it can build a unique culture where people might experience more meaningful connections, improve their personal standing and are able to look beyond differences.

It is declared that the data that support the findings of this study are not openly available because of the promise of confidentiality and anonymity with the respondents and are available from the corresponding author upon reasonable request. There is no conflict of interest. Informed consent was taken from the respondents

References

Adler, M.G. and Fagley, N.S. (2005), "Appreciation: individual differences in finding value and meaning as a unique predictor of subjective well-being", *Journal of Personality*, Vol. 73 No. 1, pp. 79-114, doi: 10.1111/j.1467-6494.2004.00305.x.

AkhavanSarraf, A.R., Abzari, M., Nasr Isfahani, A. and Fathi, S. (2017), "Generational differences in job engagement: a case study of an industrial organization in Iran", *Industrial and Commercial Training*, Vol. 49 No. 3, pp. 106-115, doi: 10.1108/ICT-10-2016-0068.

Alferjany, M.A.O.A. and Alias, R.B. (2020), "Generational differences in values and attitudes within workplace", *Psychology and Education Journal*, Vol. 57 No. 9, pp. 1496-1503, doi: 10.17762/pae.v57i9.489.

Algoe, S.B. (2012), "Find, remind, and bind: the functions of gratitude in everyday relationships", *Social and Personality Psychology Compass*, Vol. 6 No. 6, pp. 455-469, doi: 10.1111/j.1751-9004.2012.00439.x.

Algoe, S.B., Haidt, J. and Gable, S.L. (2008), "Beyond reciprocity: gratitude and relationships in everyday life", *Emotion*, Vol. 8 No. 3, pp. 425-429, doi: 10.1037/1528-3542.8.3.425.

Andert, D. (2011), "Alternating leadership as a proactive organizational intervention: addressing the needs of the baby boomers, generation Xers and millennials", *Journal of Leadership, Accountability, and Ethics*, Vol. 8 No. 4, pp. 67-83.

Becton, J.B., Walker, H.J. and Jones-Farmer, A. (2014), "Generational differences in workplace behavior", *Journal of Applied Social Psychology*, Vol. 44 No. 3, pp. 175-189, doi: 10.1111/jasp.12208.

Bentler, P.M. (1990), "Comparative fit indexes in structural models", *Psychological Bulletin*, Vol. 107 No. 2, pp. 238-246.

Bentler, P.M. and Chou, C.P. (1987), "Practical issues in structural modeling", *Sociological Methods & Research*, Vol. 16 No. 1, pp. 78-117, doi: 10.1177/0049124187016001004.

Bono, G. and Sender, J.T. (2018), "How gratitude connects humans to the best in themselves and in others", *Research in Human Development*, Vol. 15 Nos 3/4, pp. 224-237, doi: 10.1080/15427609.2018.1499350.

Brooks, A.W., Schroeder, J., Risen, J.L., Gino, F., Galinsky, A.D., Norton, M.I. and Schweitzer, M.E. (2016), "Don't stop believing: rituals improve performance by decreasing anxiety", *Organizational Behavior and Human Decision Processes*, Vol. 137, pp. 71-85, doi: 10.1016/j.obhdp.2016.07.004.

Campbell, S.M., Twenge, J.M. and Campbell, W.K. (2017), "Fuzzy but useful constructs: making sense of the differences between generations", *Work, Aging and Retirement*, Vol. 3 No. 2, pp. 130-139, doi: 10.1093/workar/wax001.

Campbell, W.K., Campbell, S.M., Siedor, L.E. and Twenge, J.M. (2015), "Generational differences are real and useful", *Industrial and Organizational Psychology*, Cambridge University Press, Vol. 8 No. 3, pp. 324-331, doi: 10.1017/iop.2015.43.

Chaudhuri, S., Park, S. and Kim, S. (2019), "The changing landscape of women's leadership in India and Korea from cultural and generational perspectives", *Human Resource Development Review*, Vol. 18 No. 1, pp. 16-46, doi: 10.1177/1534484318809753.

Chirico, A. and Yaden, D.B. (2018), "Awe: a self-transcendent and sometimes transformative emotion", *The Function of Emotions*, Springer, Cham, pp. 221-233, doi: 10.1007/978-3-319-77619-4_11.

Choi, S. and Rainey, H.G. (2010), "Managing diversity in U.S. federal agencies: effects of diversity and diversity management on employee perceptions of organizational performance", *Public Administration Review*, Vol. 70 No. 1, pp. 109-121, doi: 10.1111/j.1540-6210.2009.02115.x.

Choi, S. and Rainey, H.G. (2013), "Organizational fairness and diversity management in public organizations", *Review of Public Personnel Administration*, Vol. 34 No. 4, pp. 307-331, doi: 10.1177/0734371x13486489.

Chopik, W.J., Newton, N.J., Ryan, L.H., Kashdan, T.B. and Jarden, A.J. (2019), "Gratitude across the life span: age differences and links to subjective well-being", *The Journal of Positive Psychology*, Vol. 14 No. 3, pp. 292-302, doi: 10.1080/17439760.2017.1414296.

Cogin, J. (2012), "Are generational differences in work values fact or fiction? Multi-country evidence and implications", *International Journal of Human Resource Management*, Vol. 23 No. 11, pp. 2268-2294, doi: 10.1080/09585192.2011.610967.

Costanza, D.P., Badger, J.M., Fraser, R.L., Severt, J.B. and Gade, P.A. (2012), "Generational differences in work-related attitudes: a meta-analysis", *Journal of Business and Psychology*, Vol. 27 No. 4, pp. 375-394, doi: 10.1007/s10869-012-9259-4.

Cucina, J.M., Byle, K.A., Martin, N.R., Peyton, S.T. and Gast, I.F. (2018), "Generational differences in workplace attitudes and job satisfaction: lack of sizable differences across cohorts", *Journal of Managerial Psychology*, Vol. 33 No. 3, pp. 246-264, doi: 10.1108/JMP-03-2017-0115.

D'Amato, A. and Herzfeldt, R. (2008), "Learning orientation, organizational commitment and talent retention across generations", *Journal of Managerial Psychology*, Vol. 23 No. 8, pp. 929-953, doi: 10.1108/02683940810904402.

Deal, J.J., Stawiski, S., Graves, L., Gentry, W.A., Weber, T.J. and Ruderman, M. (2013), "Motivation at work: which matters more, generation or managerial level? ", *Consulting Psychology Journal: Practice and Research*, Vol. 65 No. 1, pp. 1-16, doi: 10.1037/a0032693.

Dries, N., Pepermans, R. and de Kerpel, E. (2008), "Exploring four generations' beliefs about career", *Journal of Managerial Psychology*, Vol. 23 No. 8, pp. 907-928, doi: 10.1108/02683940810904394.

Emmons, R.A. and McCullough, M.E. (2003), "Counting blessings versus burdens: an experimental investigation of gratitude and subjective well-being in daily life", *Journal of Personality and Social Psychology*, Vol. 84 No. 2, pp. 377-389.

Erickson, T. (2009), "Generational differences between India and the U.S", *Harvard Business Review*, available at: https://hbr.org/2009/02/global-generations-focus-on-in

Fagley, N.S. (2018), "Appreciation (including gratitude) and affective well-being: appreciation predicts positive and negative affect above the big five personality factors and demographics", *SAGE Open*, Vol. 8 No. 4, doi: 10.1177/2158244018818621.

Fagley, N.S. and Adler, M.G. (2012), "Appreciation: a spiritual path to finding value and meaning in the workplace", *Taylor & Francis*, Vol. 9 No. 2, p. 1, doi: 10.1080/14766086.2012.688621.

Fan, Y., Chen, J., Shirkey, G., John, R., Wu, S.R., Park, H. and Shao, C. (2016), "Applications of structural equation modeling (SEM) in ecological studies: an updated review", *Ecological Processes*, Vol. 5 No. 1, pp. 1-12, doi: 10.1186/s13717-016-0063-3.

Fehr, R., Fulmer, A., Awtrey, E. and Miller, J.A. (2017), "The grateful workplace: a multilevel model of gratitude in organizations", *Academy of Management Review*, Vol. 42 No. 2, pp. 361-381, doi: 10.5465/amr.2014.0374.

Felsman, P., Verduyn, P., Ayduk, O. and Kross, E. (2017), "Being present: focusing on the present predicts improvements in life satisfaction but not happiness", *Emotion*, Vol. 17 No. 7, pp. 1047-1051, doi: 10.1037/emo0000333.

Fornell, C. and Larcker, D.F. (1981), "Evaluating structural equation models with unobservable variables and measurement error", *Journal of Marketing Research*, Vol. 18 No. 1, pp. 39-50.

Garg, N. and Mehak, (2021), "Gratitude resentment and appreciation scale (GRAT-16): analyzing psychometrics properties in the Indian context", *Current Psychology*, doi: 10.1007/s12144-021-02169-6.

Garg, N., Punia, B., Kakkar, V. and Kumari, S. (2021), "Homesickness: a comparative study of defense and civilian employees and individual differences", *International Journal of Organizational Analysis*, doi: 10.1108/ijoa-07-2020-2305.

Ghosh, R. and Chaudhuri, S. (2009), "Inter-generational differences in individualism/collectivism orientations: implications for outlook towards HRD/HRM practices in India and the United States", *New Horizons in Adult Education and Human Resource Development*, Vol. 23 No. 4, pp. 5-21, doi: 10.1002/nha3.10356.

Glomb, T.M., Duffy, M.K., Bono, J.E. and Yang, T. (2011), "Mindfulness at work", *Research in Personnel and Human Resources Management*, Emerald Group Publishing Limited, Bingley, pp. 115-157, doi: 10.1108/s0742-7301(2011)0000030005.

Gursoy, D., Maier, T.A. and Chi, C.G. (2008), "Generational differences: an examination of work values and generational gaps in the hospitality workforce", *International Journal of Hospitality Management*, Vol. 27 No. 3, pp. 448-458, doi: 10.1016/j.ijhm.2007.11.002.

Hair, J.E. Jr., Anderson, R.E., Tatham, R.L. and Black, W.C. (1998), *Multivariate Data Analysis*, Prentice-Hall, Upper Saddle River, NJ.

Hlava, P. and Elfers, J. (2014), "The lived experience of gratitude", *Journal of Humanistic Psychology*, Vol. 54 No. 4, pp. 434-455, doi: 10.1177/0022167813508605.

Hobson, N.M., Schroeder, J., Risen, J.L., Xygalatas, D. and Inzlicht, M. (2017), "The psychology of rituals: an integrative review and process-based framework", *Personality and Social Psychology Review*, Vol. 22 No. 3, pp. 260-284, doi: 10.1177/1088868317734944.

Joshi, A., Dencker, J.C. and Franz, G. (2011), "Generations in organizations", *Research in Organizational Behavior*, Elsevier, Vol. 31, pp. 177-205.

Joshi, A., Dencker, J.C., Franz, G. and Martocchio, J.J. (2010), "Unpacking generational identities in organizations", *Academy of Management Review*, Vol. 35 No. 3, pp. 392-414, doi: 10.5465/amr.35.3.zok392.

Ju, B. and Li, J. (2019), "Exploring the impact of training, job tenure, and education-job and skills-job matches on employee turnover intention", *European Journal of Training and Development*, Vol. 43 Nos 3/4, pp. 214-231, doi: 10.1108/ejtd-05-2018-0045.

Kaplan, S., Bradley-Geist, J.C., Ahmad, A., Anderson, A., Hargrove, A.K. and Lindsey, A. (2014), "A test of two positive psychology interventions to increase employee well-being", *Journal of Business and Psychology*, Vol. 29 No. 3, pp. 367-380, doi: 10.1007/s10869-013-9319-4.

Keltner, D. and Haidt, J. (2003), "Approaching awe, a moral, spiritual, and aesthetic emotion", *Cognition and Emotion*, Vol. 17 No. 2, pp. 297-314, doi: 10.1080/02699930302297.

Kiken, L.G., Lundberg, K.B. and Fredrickson, B.L. (2017), "Being present and enjoying it: dispositional mindfulness and savoring the moment are distinct, interactive predictors of positive emotions and psychological health", *Mindfulness*, Vol. 8 No. 5, pp. 1280-1290, doi: 10.1007/s12671-017-0704-3.

Kline, R.B. (2009), *Becoming a Behavioural Science Researcher: A Guide to Producing Research That Matters*, Guildford Press, New York, NY, (Chapter 7, Measurement).

Kowske, B.J., Rasch, R. and Wiley, J. (2010), "Millennials' (lack of) attitude problem: an empirical examination of generational effects on work attitudes", *Journal of Business and Psychology*, Vol. 25 No. 2, pp. 265-279, doi: 10.1007/s10869-010-9171-8.

Kubacka, K.E., Finkenauer, C., Rusbult, C.E. and Keijsers, L. (2011), "Maintaining close relationships: gratitude as a motivator and a detector of maintenance behavior", *Personality and Social Psychology Bulletin*, Vol. 37 No. 10, pp. 1362-1375, doi: 10.1177/0146167211412196.

Kupperschmidt, B.R. (2000), "Multigeneration employees: strategies for effective management", *The Health Care Manager*, Vol. 19 No. 1, pp. 65-76, doi: 10.1097/00126450-200019010-00011.

Lambert, N.M., Graham, S.M. and Fincham, F.D. (2009a), "A prototype analysis of gratitude: varieties of gratitude experiences", *Personality and Social Psychology Bulletin*, Vol. 35 No. 9, pp. 1193-1207, doi: 10.1177/0146167209338071.

Lambert, N.M., Graham, S.M., Fincham, F.D. and Stillman, T.F. (2009b), "A changed perspective: how gratitude can affect sense of coherence through positive reframing", *The Journal of Positive Psychology*, Vol. 4 No. 6, pp. 461-470, doi: 10.1080/17439760903157182.

Lang, C. (2020), "Workforce diversity policies in practice: drivers and barriers in local administrations", *Ethnic and Racial Studies*, Vol. 43 No. 11, pp. 1961-1980, doi: 10.1080/01419870.2020.1754444.

Lester, S.W., Standifer, R.L., Schultz, N.J. and Windsor, J.M. (2012), "Actual versus perceived generational differences at work", *Journal of Leadership & Organizational Studies*, Vol. 19 No. 3, pp. 341-354, doi: 10.1177/1548051812442747.

Li, W., Wang, X., Haque, M.J., Shafique, M.N. and Nawaz, M.Z. (2020), "Impact of workforce diversity management on employees' outcomes: testing the mediating role of a person's job match", *SAGE Open*, Vol. 10 No. 1, doi: 10.1177/2158244020903402.

Lim, Y.-J. (2017), "Relationship between positive mental health and appreciation in Korean individuals", *International Journal of Psychology*, Vol. 52 No. 3, pp. 220-226, doi: 10.1002/ijop.12220.

Lin, C.C. (2015), "Impact of gratitude on resource development and emotional well-being", *Social Behavior and Personality: An International Journal*, Vol. 43 No. 3, pp. 493-504, doi: 10.2224/sbp.2015.43.3.493.

Lyons, S. and Kuron, L. (2014), "Generational differences in the workplace: a review of the evidence and directions for future research", *Journal of Organizational Behavior*, Vol. 35 No. S1, pp. S139-S157, doi: 10.1002/job.1913.

Ma, L.K., Tunney, R.J. and Ferguson, E. (2017), "Does gratitude enhance prosociality?: A meta-analytic review", *Psychological Bulletin*, Vol. 143 No. 6, pp. 601-635, doi: 10.1037/bul0000103.

Macky, K., Gardner, D. and Forsyth, S. (2008), "Generational differences at work: introduction and overview", *Journal of Managerial Psychology*, Vol. 23 No. 8, pp. 857-861, doi: 10.1108/02683940810904358.

Manela, T. (2016), "Gratitude and appreciation", *American Philosophical Quarterly*, Vol. 53 No. 3, pp. 281-294.

Manela, T. (2019), "The virtue of gratitude and its associated vices", *The Moral Psychology of Gratitude*, Rowman & Littlefield, London.

Manoharan, A. and Singal, M. (2017), "A systematic literature review of research on diversity and diversity management in the hospitality literature", *International Journal of Hospitality Management*, Vol. 66, pp. 77-91, doi: 10.1016/j.ijhm.2017.07.002.

McConnell, T. (2020), "Inapt gratitude: against expansionist views", *Zeitschrift Für Ethik Und Moralphilosophie*, Vol. 4 No. 1, pp. 1-18, doi: 10.1007/s42048-020-00086-0.

McMurray, A.J. and Simmers, C.A. (2020), "The impact of generational diversity on spirituality and religion in the workplace", *Vision: The Journal of Business Perspective*, Vol. 24 No. 1, pp. 70-80, doi: 10.1177/0972262919884841.

Morgan, B., Gulliford, L. and Kristjánsson, K. (2014), "Gratitude in the UK: a new prototype analysis and a cross-cultural comparison", *The Journal of Positive Psychology*, Vol. 9 No. 4, pp. 281-294, doi: 10.1080/17439760.2014.898321.

Nicuă, E.G. and Constantin, T. (2021), "Take nothing for granted: downward social comparison and counterfactual thinking increase adolescents' state gratitude for the little things in life", *Journal of Happiness Studies*, Vol. 22 No. 8, pp. 3543-3570, doi: 10.1007/s10902-021-00382-5.

Pandey, A. and Navare, A.V. (2018), "Paths of yoga: perspective for workplace spirituality", *The Palgrave Handbook of Workplace Spirituality and Fulfillment*, Palgrave Macmillan, Cham, pp. 101-126, doi: 10.1007/978-3-319-62163-0_4.

Parry, E. and Urwin, P. (2011), "Generational differences in work values: a review of theory and evidence", *International Journal of Management Reviews*, Vol. 13 No. 1, pp. 79-96, doi: 10.1111/j.1468-2370.2010.00285.x.

Piff, P.K., Dietze, P., Feinberg, M., Stancato, D.M. and Keltner, D. (2015), "Sublime sociality: how awe promotes prosocial behavior through the small self", *Journal of Personality and Social Psychology*, Vol. 108 No. 6, pp. 883-899.

Pink-Harper, S.A. and Rauhaus, B. (2017), "Examining the impact of federal employee wellness programs and employee resilience in the federal workplace", *Journal of Health and Human Services Administration*, Vol. 40 No. 3, pp. 353-387.

Popovich, P.A. (2014), "Cultivating gratitude after loss: a cross-cultural, mixed-methods research approach", Doctoral diss., Sofia University.

Rank, S. and Contreras, F. (2021), "Do millennials pay attention to corporate social responsibility in comparison to previous generations? Are they motivated to lead in times of transformation? A qualitative review of generations, CSR and work motivation", *International Journal of Corporate Social Responsibility*, Vol. 6 No. 1, pp. 1-13, doi: 10.1186/s40991-020-00058-y.

Rudolph, C. and Zacher, H. (2018), "The kids are alright: taking stock of generational differences at work", *The Industrial-Organizational Psychologist*, Vol. 55, pp. 1-7, doi: 10.31234/osf.io/wbsh5.

Rudolph, C.W., Rauvola, R.S., Costanza, D.P. and Zacher, H. (2020), "Generations and generational differences: debunking myths in organizational science and practice and paving new paths forward", *Journal of Business and Psychology*, Vol. 36 No. 6, pp. 1-23, doi: 10.1007/s10869-020-09715-2.

Rusk, R.D., Vella-Brodrick, D.A. and Waters, L. (2015), "Gratitude or gratefulness? A conceptual review and proposal of the system of appreciative functioning", *Journal of Happiness Studies*, Vol. 17 No. 5, pp. 2191-2212, doi: 10.1007/s10902-015-9675-z.

Sansone, R.A. and Sansone, L.A. (2010), "Gratitude and well being: the benefits of appreciation", *Psychiatry (Edgemont)*, Vol. 7 No. 11, pp. 18-22.

Schullery, N.M. (2013), "Workplace engagement and generational differences in values", *Business Communication Quarterly*, Vol. 76 No. 2, pp. 252-265, doi: 10.1177/1080569913476543.

Şenyuva, E. (2018), "Intergenerational differences in the personal and professional values of nurses", *Nursing Ethics*, Vol. 25 No. 7, pp. 939-950, doi: 10.1177/0969733018784688.

Simons, M., Lataster, J., Peeters, S., Reijnders, J., Janssens, M. and Jacobs, N. (2019), "Sense of abundance is associated with momentary positive and negative affect: an experience sampling study of trait gratitude in daily life", *Journal of Happiness Studies*, Vol. 21 No. 6, pp. 2229-2236, doi: 10.1007/s10902-019-00181-z.

Sommers, S. and Kosmitzki, C. (1988), "Emotion and social context: an American-German comparison", *British Journal of Social Psychology*, Vol. 27 No. 1, pp. 35-49, doi: 10.1111/j.2044-8309.1988.tb00803.x.

Srinivasan, V. (2012), "Multi generations in the workforce: building collaboration", *IIMB Management Review*, Vol. 24 No. 1, pp. 48-66, doi: 10.1016/j.iimb.2012.01.004.

Srinivasan, V., John, D.A. and Nirmala, M.C. (2011), "Understanding multigenerations in the workforce in India", *NHRD Network Journal*, Vol. 4 No. 4, pp. 121-128, doi: 10.1177/0974173920110423.

Srinivasan, V., John, D.A. and Christine, M.N. (2014), "Generational cohorts and personal values: an exploratory study in the indian workplace", in Parry, E. (Ed.), *Generational Diversity at Work: New Research Perspective*, Routledge, New York, NY, pp. 185-206.

Steindl-Rast, D. (2004), "Gratitude as thankfulness and as gratefulness", in Emmons, R.A. and Mccullough, M.E. (Eds), *The Psychology of Gratitude*, Oxford University Press, Oxford, pp. 282-291.

Stellar, J.E., Gordon, A., Anderson, C.L., Piff, P.K., McNeil, G.D. and Keltner, D. (2018), "Awe and humility", *Journal of Personality and Social Psychology*, Vol. 114 No. 2, pp. 258-269, doi: 10.1037/pspi0000109.

Strauss, W. and Howe, N. (1991), *Generations: The History of America's Future*, William Morrow and Company, New York, NY, pp. 1584-2069.

Tamunomiebi, M.D. and John-Eke, E.C. (2020), "Workplace diversity: emerging issues in contemporary reviews", *International Journal of Academic Research in Business and Social Sciences*, Vol. 10 No. 2, doi: 10.6007/ijarbss/v10-i2/6926.

Tucker, K.L. (2007), "Getting the most out of life: an examination of appreciation, targets of appreciation, and sensitivity to reward in happier and less happy individuals", *Journal of Social and Clinical Psychology*, Vol. 26 No. 7, pp. 791-825, doi: 10.1521/jscp.2007.26.7.791.

Tudge, J.R., Freitas, L.B. and O'Brien, L.T. (2015), "The virtue of gratitude: a developmental and cultural approach", *Human Development*, Vol. 58 Nos 4/5, pp. 281-300, doi: 10.1159/000444308.

Twenge, J.M. (2010), "A review of the empirical evidence on generational differences in work attitudes", *Journal of Business and Psychology*, Springer, Vol. 25 No. 2, pp. 201-210, doi: 10.1007/s10869-010-9165-6.

Vanderschuere, M. and Birdsall, C. (2018), "Can diversity management improve job satisfaction for military veterans in the federal government?", *The American Review of Public Administration*, Vol. 49 No. 1, pp. 116-127, doi: 10.1177/0275074018783005.

Watkins, P.C. (2004), "Gratitude and subjective well-being", in Emmons, R.A. and McCullough, M.E. (Eds), *Psychology of Gratitude*, Oxford University Press, Oxford, pp. 167-195.

Watkins, P., Woodward, K., Stone, T. and Kolts, R.L. (2003), "Gratitude and happiness: development of a measure of gratitude, and relationships with subjective well-being", *Social Behavior and Personality: An International Journal*, Vol. 31 No. 5, pp. 431-451.

Weeks, K.P. and Schaffert, C. (2019), "Generational differences in definitions of meaningful work: a mixed methods study", *Journal of Business Ethics*, Vol. 156 No. 4, pp. 1045-1061, doi: 10.1007/s10551-017-3621-4.

Wong, M., Gardiner, E., Lang, W. and Coulon, L. (2008), "Generational differences in personality and motivation: do they exist and what are the implications for the workplace?", *Journal of Managerial Psychology*, Vol. 23 No. 8, pp. 878-890, doi: 10.1108/02683940810904376.

Wood, A., Joseph, S. and Linley, A. (2007), "Gratitude – parent of all virtues", *Psychologist*, Vol. 20, pp. 18-21.

Wood, A.M., Froh, J.J. and Geraghty, A.W.A. (2010), "Gratitude and well-being: a review and theoretical integration", *Clinical Psychology Review*, Vol. 30 No. 7, pp. 890-905, doi: 10.1016/j.cpr.2010.03.005.

Zopiatis, A., Krambia-Kapardis, M. and Varnavas, A. (2012), "Y-ers, x-ers and boomers: investigating the multigenerational (mis)perceptions in the hospitality workplace", *Tourism and Hospitality Research*, Vol. 12 No. 2, pp. 101-121, doi: 10.1177/1467358412466668.

Corresponding author

Manju Mahipalan can be contacted at: mmanju@nitt.edu

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

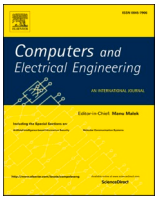
Computers and Electrical Engineering 101 (2022) 108035



Contents lists available at ScienceDirect

Computers and Electrical Engineering

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





Fetal state health monitoring using novel Enhanced Binary Bat Algorithm

Prerna Sharma, Kapil Sharma *

Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Cardiotocography
Classification
Feature selection
Evolutionary algorithm
Bat algorithm
Binary Bat Algorithm
Fetal distress

ABSTRACT

The timely assessment of hypoxic fetuses by cardiotocography is of utmost significance as deficiency of oxygen in fetuses' leads to fetal distress which can further prove to be fatal or cause neurological diseases. This paper puts forward an Enhanced Binary Bat algorithm, the altered form of Binary Bat Algorithm for the multi-classification problem of Cardiotocography. Subset of optimal and relevant features is selected using the optimized and Enhanced BBA algorithm from cardiotocography dataset. Features selected by various evolutionary algorithms EBBA (Enhanced Binary Bat algorithm), qGWO (quantum Grey Wolf Optimization), Genetic algorithm are 11, 15, 12 simultaneously. EBBA efficiently selects most reduced set of features. The proposed EBBA can be used in feature selection and classification of cardiotocography dataset under different fetus state i.e. normal, suspect and pathologic with an accuracy of 96.21% using machine learning classifier i.e. random forest.

1. Introduction

These days' uncertainty and prevalent conditions of pandemic has engulfed and affected the world to an immense extent. Health has become top priority worldwide. Pregnant ladies were considered to be at higher risk to this pandemic situation. Otherwise also pregnancy may be complicated due to various medical problems associated with mother's health that could affect fetal well-being in turn. Risk to the fetus tends to increase with mother facing medical conditions like hypertension, preeclampsia, maternal diabetes and thyroid disease. There are many other circumstances that escalate level of risk to fetus during pregnancy such as vaginal bleeding, intra uterine growth restriction, extended pregnancy, placenta insufficiency, lessened fetal movements, multiples pregnancy etc. These complicated condition leads to the need of an additional means to gauge the fetal well-being.

Cardiotocography is a process in which fetal heart rate and uterine contractions are monitored incessantly during pregnancy by employing ultrasound transducer upon the abdomen of the mother [1]. At times, the process of cardiotocography is well known as Electronic Fetal Monitoring (EFM). The purpose is to figure out the babies to whom supply of oxygen may possibly fall short; such a condition suffered by a fetus is called fetal hypoxia, which can have severe and prolonged effect as there is a risk associated with the baby of still birth, born with disabilities. It is an indicative tool that is prevalently used to ascertain the fetal distress in the course of antepartum and intra-partum phase. With the help of readings and evaluations, one can get hold of Anabolic Acidosis and Hypoxic injury [2]. Cardiotocograph or more generally known as Electronic Fetal Monitor (EFM) is a machine used to keep record of heartbeat of fetus and the uterine contractions at the time of pregnancy. This process is popularly and formally known as Cardiotocography

* Corresponding author.
E-mail address: kapil@ieee.org (K. Sharma).

https://doi.org/10.1016/j.compeleceng.2022.108035
Received 30 January 2022; Received in revised form 19 April 2022; Accepted 21 April 2022
Available online 19 May 2022
0045-7906/© 2022 Published by Elsevier Ltd.

(CTG).The cardiotocographic data aids in classification of pathological state of the fetus as per normal indicating healthy condition, suspect with appropriate action required determining the extent of fetal distress or pathologic with speedy action required as fetal distress can affect immensely. It helps the obstetrician to foresee problems associated and obstruct the harm that can be caused to the fetus. During delivery, Hypoxic fetus is vulnerable to temporary impairment or death. Unsuitable treatments and incorrect diagnosis of Fetal Heart Rate pattern recordings results in more than half of the caused fatalities [3–5]. There is reduction in number of neonatal seizures with incessant cardiotocography during labor [6].

In terms of reliability External FHR outpaces Internal FHR as it also takes other parameters into account like recording maternal heart rate, loss of signal etc. likely during second stage of labour [7]. Before the CTG recording starts, it is advised to check the maternal heart rate separately in order to avoid the confusion of maternal heart rate and fetus heart rate [8]. During labour if there is a doubt of recording maternal heart rate instead of fetus, immediate internal monitoring is required with the help of fetal scalp electrode(FSE), if required. Obstetricians can make use of supplementary decision support system based on machine learning during Cardiotocography trace interpretation as an impartial gauge tool to enhance predictive competence and greatly reduces negative outcomes [9]. Generally growth restricted or a preterm fetus has low reserves and compensatory mechanisms in comparison to full term grown fetus [10].

Nature-Inspired Algorithms (NIAs) are fundamentally the algorithms for whom the nature becomes source of inspiration and motivation, nature is itself intellectualized with significant spread of algorithms and solutions buried in it; if we closely observe them they could be of great help in figuring out our problems [11]. Nature-inspired computing has got together numerous poles apart disciplines of biology, mathematics and computer science. These algorithms are based on the technique of bio-inspired computing which is pretty much progressive to new and different challenging techniques which are grounded on the principle of natures' evolution [12].

Evolutionary Algorithms/Bio-inspired emulate the evolutionary act of creatures set up in nature. The search algorithms initiate with forming population using randomly generated solutions that tends to evolve throughout succeeding generations. Medical data when treated with various computational intelligence methods helps in getting better grip of efficiency in terms of speed and precision detection. In the medical field nature inspired techniques are categorized under two different heads by using single and hybrid methods and nature inspired algorithms dealing with medical data got exposure of various state- of-the- art methodologies [13], these days' applications of science and engineering likewise decision making, objective optimization and processing of the information etc. are making use of various techniques of nature inspired algorithm.

Owing to redundancy of features and the burden of dimensionality, the learning engine reserves a considerable toll on time and the efficiency of the model drops [14]. Application classification analysis to support medical diagnostic decisions, improve the quality of patients' care etc. A subset of data sets containing vast volumes of data contained in medical records is chosen for instruction. If the training data set contains unrelated and inapt characteristics, the classification analysis could yield results that are less in precision and not as much as comprehensible as needed. The selection of the feature sub-sets is an extremely important pre-processing stage in the field of data extraction [15]. The enhancement of the health database increases the diagnostic efficiency. Some extra attributes can increase the computation time that has an effect on the accuracy of the diagnosis. Any data in the data collection may not be useful for diagnosis and may therefore be discarded prior to learning.

Subrota Mazumdar et.al. proposed a technique that monitors fetal heart rate using Artificial Neural Network [16]. Vinayaka Nagendra et.al. discussed the significance of machine learning for backing obstetricians in pathologic suspect cases [17], Huang and Hsu [18] evaluated the fetal distress on CTG data using various techniques of ANN, discriminant analysis(DA) and decision tree with the accuracies of 97.78%, 82.1%,86.36%, respectively. Anish Batra et.al. evaluated the fetal distress using combination of various machine learning algorithms [19]. Chinnasamy et al. [20] proposed classification of CTG data using supervised ANN.

Yang [21] proposed an exclusive nature-inspired metaheuristic algorithm namely Bat algorithm which tends to mimic the echolocation technique used by bats to identify an obstacle or a prey. Mirjalili et al. [22] proposed an altered version of existing Bat algorithm for binary problems. Considerable reduction in computational time and cost is desired Nakamura et al. mentioned that the Binary Bat Algorithm feature selection approach has outdone various popular swarm-based techniques in numerous public datasets. Yang et al. [23] proposed a modified version algorithm namely MBAFS and also recorded its comparison with BA and BPSO based algorithms on numerous standard datasets.Similar application was also presented by Gupta and Ahlawat [24]. A number of optimization algorithms which are Bio-Inspired or Nature Inspired like Firefly algorithm (FA), Cuckoo Search(CSO), Particle Swarm Optimization(PSO), Cuttlefish algorithm (CFO) striving local optima was presented by Xin-She Yang in his book [25].

This paper proposes an optimized form of customary Binary Bat algorithm namely Enhanced Binary Bat Algorithm (EBBA). The algorithm proficiently combines the feature selection with the echolocation principle observed by bats for the cardiotocograph dataset. Further the algorithm makes use of several machine learning classifiers to report the accuracies of various nature inspired algorithms. The EBBA algorithm has been applied on the CTG (Cardiotocography) dataset is accessible at UCI Machine Learning Repository (https://archive.ics.uci.edu/ml/machine-learning-databases/00193/). EBBA efficiently reduces the number of features i.e. selects the most distinguishing features with an average accuracy of 96.21%. Hence, it proves to be effective in the classification of fetal state under normal, suspect and pathologic label.

The rest of the paper is segregated under various sections. The background study is staged in Section 2. The detailed explanation of the methodology used is placed under Section 3. Sections 4 exhibits results of novel and other algorithms used and its comparison with contemporary algorithms. The paper is concluded and future scope is highlighted in Section 5.

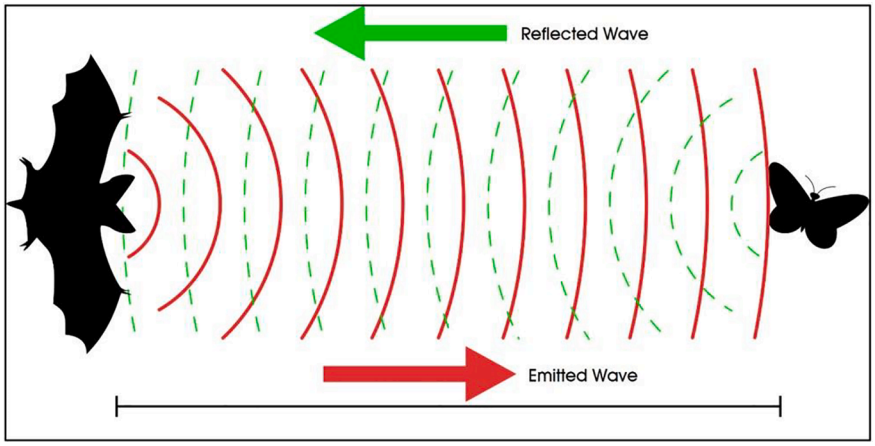


Fig. 1. Echolocation in Micro Bats.

2. Background

2.1. Bat algorithm

This algorithm is inspired by captivating group of mammals called bats. They possess the peculiar ability of echolocation as shown in Fig. 1 which they extensively use to detect an obstacle in their way during movement or track prey/food because of which they have drawn the attention of researchers from numerous fields. The phenomenon used by them is akin to sonar. Micro bats play a fundamental part in developing the significant attributes of this algorithm. They emit short pulse of sound which is also loud in nature, an echo returns as this wave knocks an object after a slice of time [25]. This in turn helps the bats to learn their farness from the object. These loud ultrasound generated tends to fluctuate around a fixed frequency which helps them to catch their prey and escape the obstacle as they are able to discriminate amongst them. A novel and interesting metaheuristics optimization technique has been developed by Yang [14] on the basis of behavior of group of bats approaching obstacles, food/prey.

For each bat bi, its initial position χ_i, velocity v_i and frequency f_i are initialized with varying wavelength and loudness i.e. λ *and*A_0 to search for prey. T depicts the maxi mum number of iterations whereas t depicts each time step. As they approach the target, they tend to emit pulse whose wavelength (or frequency) is habitually adjusted, the rate of pulse emission ranges between values of 0 and 1 which greatly depends on its closeness to the object. The value of loudness varies from A_0 to A_{min} i.e. from a large positive value to a minimum constant value. Eqs. (1)–(3) mentioned below keep record of velocity and position changes by the travelling of virtual bats.

$$f_i = f_{\min} + (f_{\min} - f_{\max})\beta \tag{1}$$

$$v_i^j\ (t) = v_i^j\ (t\ -1)\ + \left[\widehat{x}j - \chi_j^i(t-1)\right] fi \tag{2}$$

$$x^j{}_i(t) = x^j{}_i(t-1) + v^j{}_i(t) \tag{3}$$

Here,

β-> randomly generated number β∈ [0, 1].

χ_i^j -> value of decision variable j for bat i at time step t.

Bat movements speed and reach is regulated by the value of f_i obtained by Eq. (1). The value of x^j variable depicts global best location solution for the decision variable j at that moment by comparing all the solutions of m bats available. Yang has recommended taking up random walks to get better variability of the likely solutions. Selection of one appropriate solution is made out of the current best solutions available and then to obtain a new solution for each and every bat that satisfies line 8 of algorithm 1 random walk is applied as given in Eq. (4).

$$x_{new} = x_{old} + \in \overline{A}(t) \tag{4}$$

(*t*) stands for the average value of loudness of all the bats at time *t*, ∈ *varies in the range of* [−1, 1] attempts to the direction and strength of the random walk. The values are brought up-to-date for loudness A_i *and* emission pulse rate r_i *for* each and every iteration as mentioned below in Eqs. (5) and (6):

$$(t+1) = (t) \tag{5}$$

and

$$(t+1) = (0)[1 - exp(-\gamma t)] \tag{6}$$

Algorithm 1
Bat Algorithm

1.	Initially set the Bat population as b_i,where [i= 1 to m]
2.	For each bat b_i there is a position x_i, velocity v_i and pulse frequency f_i where [i= 1 to m]
3.	Initialize pulse rates as pulse_rate r_i, the loudness as A_i
4.	Initialize each iteration as iter in T where T is maximum number of iterations
5.	While iter<T, Repeat steps 6–18
6.	For bat b_i, Repeat steps 7–15
7.	Generate new solutions using Eqs. (1)–(3)
8.	If (rand > r_i)
9.	Choose a solution from the best solutions
10.	Produce a local solution around the best solution
11.	End if
12.	If (rand < A_i) & (f (x_i) < f (CurrentBest))
13.	Increase pulse rate r_i and reduce loudness of bat Ai
14.	End if
15.	End for
16.	Rank the bats and find CurrentBest
17.	End if
18.	End while

Algorithm 2
Enhanced Binary Bat Algorithm

Set population size size_of_pop, pulse emission rate constant pulse_rate_constant number of iterations iter, loudness constant loud_constant, total features n, bat position pos_of_bat, bat velocity vel_of_bat, bat loudness loud_of_bat, pulse emission rate pulse_rate, fitness fitness, global fitness global_fitness
for each model do:
test_train_split(ratio = 70:30) . Sets should have only those features where pos_of_bat[i] = 1
Fitness function is defined as the accuracy of the machine learning model. Accuracy is updated to acc[i]
Set **random = [0,1]**
if ((loud_of_bat[i] >= random) and (acc >= fitness[i]):
fitness[i] = acc[i]
Update pulse_rate[i] & loud_of_bat[i] using EQUATIONS
end if
end for
Set index_pos and max_fitness
if (max_fitness > global_fitness):
global_fitness = max_fitness
global_pos = index_pos
end if
Attributes Selected = Numeber of 1's in global_pos
for each model do:
Beta = random = e = [0,1]
if (random > r[i]):
find pos_of_bat[i] using EQUATIONS
end if
if ((loud_of_bat[i] > random) **and** (global_fitness > fitness[i])):
Update freq_of_bat[i], vel_of_bat[i], pos_of_bat[i] using EQUATIONS
Get new pos_of_bat[i] using EQUATIONS
end if
end for

where α, γ are ad-hoc constants. Initially in the algorithm the values for emission rate (0) and the loudness (0) are chosen randomly most of the time. Generally, (0) ∈ [1,2] and (0) ∈ [0, 1] [12].

The following Algorithm 1 depicts the aforesaid BAT algorithm in detail.

2.2. Feature selection and binary bat algorithm

Bat algorithm application to continuous valued problems was exhibited by Yang [21]. In this the search space contained continuous valued positions that a bat can acquire. However, for combinatorial and discrete problems Nakamura et al. came up with the new variant of bat algorithm called Binary Bat Algorithm. On the other hand, for feature selection the search space is exhibited in the form of Boolean lattice of n dimension where the movement of bat tends to happen across the corners and nodes of the lattice. Since the problem addresses whether a feature is selected represented by value 1 or not by value 0. Eq. (7) depicts bat's position which is controlled by a sigmoid function and the position of the bat is restricted using Eq. (8). The sigmoid function is used to restrict the bats position to only 0 and 1 i.e. binary values in the binary version of the bat algorithm [25].

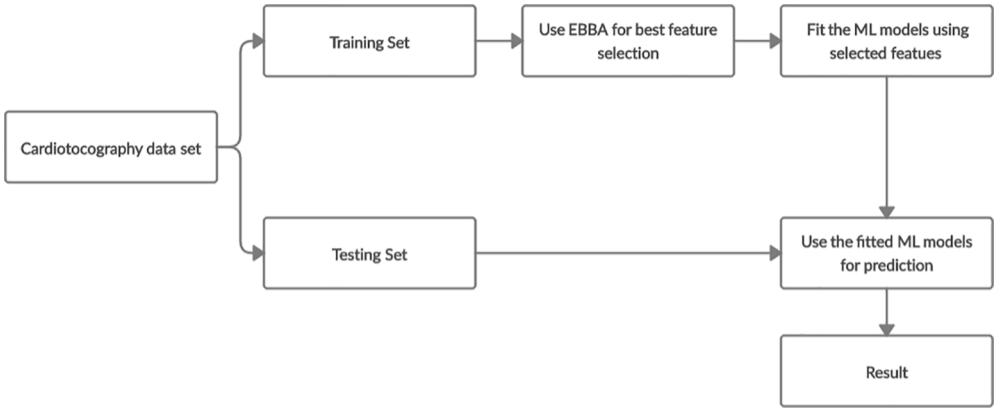


Fig. 2. Flow graph of EBBA.

$$S(V_i) = \frac{1}{1 + e^{-V_i}} \tag{7}$$

$$\chi_i^j = \ 1 \text{ if } S\left(v_j^i\right) > \ \sigma \ 0 \text{ Otherwise} \tag{8}$$

in which σ ∼ $U(0, 1)$. Hence Eq. (8) makes 0, 1 i.e. binary values available to the Boolean lattice for bat's coordinates which signifies whether the feature is selected or not. Eq. (3) can be replaced by Eq. (8). Binary Bat algorithm is explicitly explained in the Sharma et. al. [11].

3. Methodology

3.1. Enhanced Binary Bat Algorithm (EBBA) for feature selection and classification

The EBBA algorithm is modified version of Binary Bat Algorithm. It accepts all the attributes of the given IoT dataset and returns back the optimal features. It has been applied on Cardiotocography dataset. It is a multi-classification problem in which the diagnosis is given by an obstetrician about the fetal state i.e. Normal, suspect and pathologic. The EBBA was used with three classification machine learning models KNN, DecisionTree, Random Forest. The following Algorithm 2 explains each step of EBBA precisely.

In Fig. 2 Cardiotocography dataset of 2126 fetal cardiotocograms (CTGs) features is divided into training set and testing set with the ratio of 70:30. EBBA is employed to the training set for the selection of optimal features. Machine classifiers are then applied to find out the accuracy of solutions obtained with the optimal features selected by EBBA. Result is the dataset classification with the determination of fetus state accurately. Fig. 3 comprehensively explains the working of Enhanced Binary Bat Algorithm (EBBA).

3.2. Dataset

The dataset contains measurements of Fetal Heart Rate (FHR) and uterine contraction (UC) features on cardiotocograms classified by expert obstetricians. The corresponding diagnostic features were evaluated by treating 2126 fetal cardiotocograms (CTGs).Three expert obstetricians classified CTGs with a classification label. Classification is primarily performed under two heads; one is morphologic pattern (A, B, C. ...) and the other is fetal state (N, S, P). The dataset is used for 3-class classification. Table 1 comprises of description of various Cardiotocograms attributes.

Each cardiogram is represented in the dataset by 21 features out of which 11 are obtained using electronic fetal monitoring sensors and the rest 10 are extracted using those 11 features. Table 2 exhibits various parameters of problem type and their values in cardiotocography dataset in terms of number of attributes, sample, features and also types of features in using multi-classification problem.

3.3. Implementation of EBBA

Implementation of EBBA has been performed. This section gives an insight to various implementation related information in terms of set up required to perform the experiment, parameters taken as inputs, various tuning parameters of machine learning algorithms.

3.3.1. Experimental setup

The system on which all the testing was carried out has the following configuration:

- Windows 10 version 1903

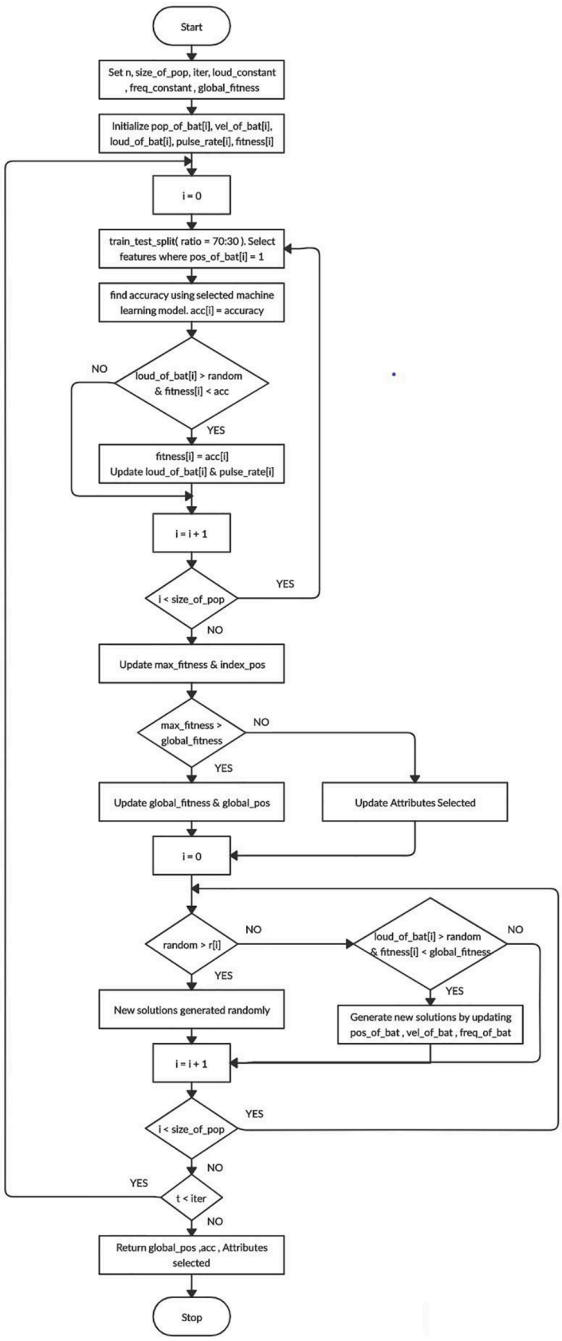


Fig. 3. Flow chart of EBBA.

- 2.30 GHz Intel Core i5 6th generation
- 8 GB 2401 MHz DDR4

Python 3.7 and it's libraries have been used.

3.3.2. Input parameters

This section talks about various input parameters taken into consideration for EBBA. The variables used in the algorithm are mentioned in Table 3.

Table 1
Description of ctg attribute.

ATTRIBUTE	DESCRIPTION
LB	FHR baseline (beats per minute)
AC	No. of accelerations per second
FM	No. of fetal movements per second
UC	No. of uterine contractions per second
DL	No. of light decelerations per second
DS	No. of severe decelerations per second
DP	No. of prolonged decelerations per second
ASTV	percentage of time with abnormal short-term variability
MSTV	mean value of short-term variability
Width	the width of FHR histogram
Min	minimum of FHR histogram
Max	Maximum of FHR histogram
Nmax	No. of histogram peaks
Nzeros	No. of histogram zeros
Mode	histogram mode
Median	histogram median
Variance	histogram variance
Tendency	histogram tendency
CLASS	FHR pattern class code (1 to 10)
NSP	fetal state class code (N = normal; S = suspect; P = Pathologic)

Table 2
Mapping of Problem type and their values in cardiotocography dataset.

Problem Type	Multi – Classification Problem
Types of features	Real
Total attributes	23
Total samples	2125
Total missing values	0

Table 3
Initialized value and description of input parameters.

Parameter	Initialized Value	Description
n	**21**	Total features
iter	**20**	Total iterations
size_of_pop	**30**	Size of population
pos_of_bat	**[0,1]**	Position of bats
vel_of_bat	**0**	Velocity of bats
loud_constant	**0.8**	Loudness constant
freq_constant	**[0,1]**	Frequency constant
pulse_rate_constant	**0.9**	Pulse rate constant
loud_of_bat	**[1,2]**	Loudness
pulse_rate	**[0,1]**	Pulse rate
fitness	**-1**	Fitness
global_fitness	**-1**	Global Fitness

Table 4
Parameters of classifiers.

Model	Tuning Parameters
Decision Tree	max_depth = 16
K-NN	p = 1,n_neighbors =1
Random Forest	n_estimators = 105

3.3.3. Machine learning models

Different machine learning models (DecisionTree, K-nearest neighbours,Random Forest) are used for classification. Table 4 gives insight to various tuning parameters associated with machine learning classifiers.

Table 5
No. of features selected using different evolutionary algorithms.

Algorithm	Features selected (out of 21)
EBBA	11
qGWO	15
Genetic	12

Table 6
Accuracy of different ML classifiers on all features and features selected by EBBA.

Model	Accuracy using all features	Accuracy using selected features
Decision Tree	92.63	94.36
K-NN	92.78	94.67
Random Forest	93.57	96.21

Table 7
Comparison of accuracy obtained by ML classifiers on the basis of features selected by different Evolutionary Algorithms.

Algorithm / ML Classifier	Decision Tree	K-NN	Random Forest
EBBA	94.36	94.67	96.21
gGWO	93.41	92.47	94.82
Genetic	94.04	92.06	94.66

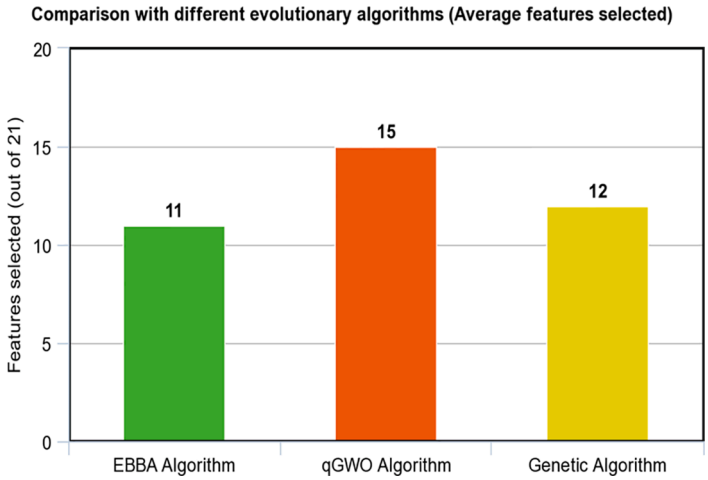


Fig. 4. Number of features selected using different Evolutionary Algorithms.

4. Results

In this section results of the EBBA algorithm on Cardiotocography dataset have been discussed. The number of optimal features selected using different evolutionary algorithms is shown in Table 5 which shows EBBA tends to select the reduced set of 11 features out of 21 whereas qGWO selects 15 and Genetic Algorithm selects 12 features. Table 6 exhibits accuracy in percentage obtained using different ML models (Decision Tree, K-NN, Random Forest) when applied to all features of the dataset and optimal features selected by proposed EBBA. Comparative study has been done with the nature inspired algorithms qGWO and Genetic Algorithm in terms of accuracy. Accuracy of different algorithms on the basis of features selected by EBBA is presented in Table 7. There is an average increase of 2.09 % in accuracy when using EBBA. The proposed algorithm outperforms the two evolutionary algorithms in terms of accuracy and minimal feature subset obtained. Average number of features selected by different evolutionary algorithms EBBA, qGWO and GA is depicted in Fig. 4. Fig. 5 presents accuracy computed in terms of percentage using different ML classifiers on the set of all features and set of features selected by EBBA. Comparative analysis of accuracy of different evolutionary algorithms and machine learning classifiers is represented in Fig. 6.

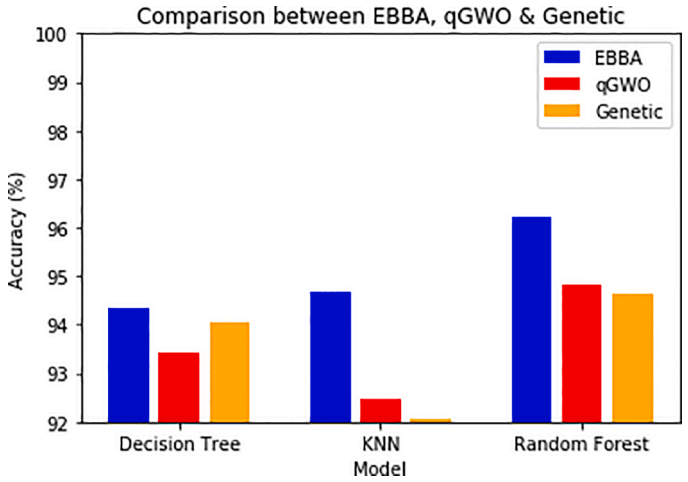


Fig. 5. Accuracy computed in percentage using different ML classifier using all features and the features selected using EBBA.

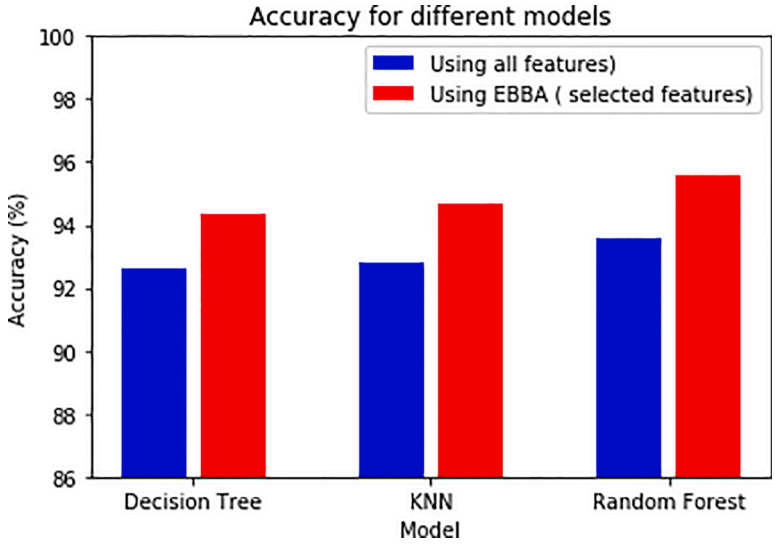


Fig. 6. Comparison between different evolutionary algorithms & ML classifiers in terms of Accuracy.

5. Conclusion and future scope

Enhanced Binary Bat Algorithm (EBBA) outperforms other metaheuristics for optimized feature selection. This novel bio-inspired optimization algorithm greatly reduces computational cost by performing dimensionality reduction with high accuracy for implementing feature selection. Section 3.1 discusses the suggested algorithm in detail. EBBA has made a selection of least possible 11 features, with the utmost average accuracy of 96.21% in comparison to average accuracies of 94.82% and 94.66% computed by customary algorithms qGWO and GA, respectively. The results depict that this novel proposed algorithm outperforms qGWO and GA with respect to accuracy measure.

Further studies can be conducted in the discipline of fetal health monitoring and diagnosing of fetal distress could be done by combining various nature inspired algorithms and various machine learning models. The hybridization of various nature inspired algorithms can give results with improved accuracies. Other techniques of fetal distress determination can also be evaluated and examined with the proposed algorithm for timely and accurate detection. Other than the cardiotocography data set EBBA can moreover be useful to arena of optimization problems applications.

Conflict of Interest

None.

Data statement

No data is available.

Declaration of Competing Interest

None.

References

[1] Miao JH, Miao KH. Cardiotocographic diagnosis of fetal health based on multiclass morphologic pattern predictions using deep learning classification. Int. J. Adv. Comput. Sci. Appl. 2018;9(5):1–11.
[2] Ramla M, Sangeetha S, Nickolas S. Fetal health state monitoring using decision tree classifier from cardiotocography measurements. In: Proceedings of the 2nd International Conference on Intelligent Computing and Control Systems (ICICCS); 2018.
[3] Marzbanrad F, Stroux L, Clifford GD. Cardiotocography and beyond: a review of onedimensional Doppler ultrasound application in fetal monitoring. Physiol. Meas. 2018;39(8):08TR01.
[4] Subasi A, Kadasa B, Kremic E. Classification of the cardiotocogram data for anticipation of fetal risks using bagging ensemble classifier. Procedia Comput. Sci. 2020;168:34–9.
[5] Nagendra V, Gude H, Sampath D, Corns S, Long S. Evaluation of support vector machines and random forest classifiers in a real-time fetal monitoring system based on cardiotocography data. In: Proceedings of the IEEE Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB); 2017.
[6] Alfirevic Z, Devane D, Gyte GM, Cuthbert A. Continuous cardiotocography (CTG) as a form of electronic fetal monitoring (EFM) for fetal assessment during labour. Cochrane Database Syst. Rev. 2017;2(5):CD006066.
[7] Pinas A, Chandraharan E. Continuous cardiotocography during labour: analysis, classification and management. Best Pract. Res. Clin. Obstet. Gynaecol. 2016; 30:33–47.
[8] Nurani R, Chandraharan E, Lowe V, Ugwumadu A, Arulkumaran S. Misidentification of maternal heart rate as fetal on cardiotocography during the second stage of labor: The role of the fetal electrocardiograph. Obstet. Gynecol. Surv. 2013;68(4):261–3.
[9] Fergus P, Selvaraj M, Chalmers C. Machine learning ensemble modelling to classify caesarean section and vaginal delivery types using Cardiotocography traces. Comput. Biol. Med. 2018;93:7–16.
[10] Afors K, Chandraharan E. Use of continuous electronic fetal monitoring in a preterm fetus: clinical dilemmas and recommendations for practice. J. Pregnancy 2011;2011:848794.
[11] Zang H, Zhang S, Hapeshi K. A review of nature-inspired algorithms. J. Bionic Eng. 2010;7(S4):S232–7.
[12] Sharma P, Sharma K. A novel quantum-inspired binary bat algorithm for leukocytes classification in blood smear. Expert Syst. 2021.
[13] Kalantari A, Kamsin A, Shamshirband S, Gani A, Alinejad-Rokny H, Chronopoulos AT. Computational intelligence approaches for classification of medical data: state-of-the-art, future challenges and research directions. Neurocomputing 2018;276:2–22.
[14] Jain I, Jain VK, Jain R. Correlation feature selection based improved-binary particle swarm optimization for gene selection and cancer classification. Appl. Soft Comput. 2018;62:203–15.
[15] M.Harb H, Desuky AS. Feature selection on classification of medical datasets based on particle swarm optimization. Int. J. Comput. Appl. 2014;104(5):14–7.
[16] Mazumdar S, Choudhary R, Swetapadma A. An innovative method for fetal health monitoring based on artificial neural network using cardiotocography measurements. In: Proceedings of the 3rd International Conference on Research in Computational Intelligence and Communication Networks (ICRCICN; 2017.
[17] Nagendra V, Corns S. Evaluation of Support Vector Machines and Random Forest Classifiers in a Real time Fetal Monitoring System Based on Cardiotocography Data. IEEE; 2017.
[18] Huang M-L, Hsu Y-Y. Fetal distress prediction using discriminant analysis, decision tree, and artificial neural network. J. Biomed. Sci. Eng. 2012;05(09):526–33.
[19] Batra A, Chandra A, Matoria V. Cardiotocography analysis using conjunction of machine learning algorithms. In: Proceedings of the International Conference on Machine Vision and Information Technology (CMVIT); 2017.
[20] Chinnasamy S, Chitradevi MCM, Geetharamani G. Classification of cardiotocogram data using neural network based machine learning technique. Int. J. Comput. Appl. 2012;47(14):19–25.
[21] Yang X-S. A New Metaheuristic Bat-Inspired Algorithm. Nature Inspired Cooperative Strategies for Optimization (NICSO 2010). Berlin, Heidelberg: Springer Berlin Heidelberg; 2010. p. 65–74.
[22] Mirjalili S, Mirjalili SM, Yang XS. Binary bat algorithm. Neural Comput. Appl. 2014;25(3–4):663–81.
[23] Yang B, Lu Y, Zhu K, Yang G, Liu J, Yin H. Feature selection based on modified bat algorithm. IEICE Trans. Inf. Syst. 2017;E100.D(8):1860–9.
[24] Gupta D, Ahlawat AK. Usability feature selection via MBBAT: a novel approach. J. Comput. Sci. 2017;23:195–203.
[25] Gupta D, Arora J, Agrawal U, Khanna A, de Albuquerque VHC. Optimized binary bat algorithm for classification of white blood cells. Measurement 2019;143: 180–90 (Lond.).

Prerna Sharma is an Assistant Professor in the Department of Computer Science and Engineering at MAIT, GGSIPU, India. She is also a doctoral researcher at Delhi Technological University (DTU), India. She has obtained her M.Tech in IT from USIT, GGSIPU India and B.Tech in CSE from GPMCE, GGSIPU, India. Her research interest includes Artificial Intelligence, Computational Intelligence Nature-Inspired Computing.

Dr. Kapil Sharma is a Professor & Head of Information Technology at Delhi Technological University Delhi, India. He received his PhD in Computer Science from Maharshi Dayanand University, India. He has obtained his B.E. and M.Tech Degrees in Computer Science from MD University and IASE University, India, respectively. His research interests include system design, pattern recognition, computer vision, soft computing.

arXiv:2205.01366v2 [cs.CL] 4 May 2022

Finding patterns in Knowledge Attribution for Transformers

Jeevesh Juneja
Delhi Technological University
creativityinczenyoga@gmail.com

Ritu Agarwal
Delhi Technological University
rituagarwal@dtu.ac.in

May 5, 2022

Abstract

We analyze the Knowledge Neurons(Dai et al., 2021) framework for the attribution of factual and relational knowledge to particular neurons in the transformer network. We use a 12-layer multi-lingual BERT model for our experiments. Our study reveals various interesting phenomena. We observe that mostly factual knowledge can be attributed to middle and higher layers of the network(≥ 6). Further analysis reveals that the middle layers($6-9$) are mostly responsible for relational information, which is further refined into actual factual knowledge or the "correct answer" in the last few layers($10-12$). Our experiments also show that the model handles prompts in different languages, but representing the same fact, similarly, providing further evidence for effectiveness of multi-lingual pre-training. Applying the attribution scheme for grammatical knowledge, we find that grammatical knowledge is far more dispersed among the neurons than factual knowledge.

1 Introduction

Geva et al. (2020) showed that the feed-forward layers of the transformer (Vaswani et al., 2017) which constitute almost two-thirds of the model's parameters operate as a key-value store, with each key correlating with textual patterns which the model has successfully learned to detect. The values corresponding to these keys are indicative of the output distributions over vocabulary of the model for that particular pattern.

Dai et al. (2021) extends this to relate "facts" (for e.g. that capital of France is Paris) to certain key-value memories, i.e., to certain neurons in the feed-forward layers of the transformer model. They conduct experiments which show that modulating these particular neurons' activations has much larger impacts on the correct output's probabilities, as compared to the case where they modulate random neurons' activations.

In the reverse direction, they also show that queries related to the fact tend to activate these particular neurons more than others. Finally, they provide methods to update and erase preexisting knowledge/facts in these neurons without any fine-tuning.

We explore their attribution strategy further, with different kinds of knowledge, at different layers of the network, proposing hypotheses for inner working of transformer language models.

2 Related Works

The concept of "memory" of a model has been one that has existed for a long time and has helped push the state-of-the-art of language models further consistently(Hopfield (1982), Hochreiter and Schmidhuber (1997)). As the models become larger and larger, they store more and more memories, but it becomes harder to learn and retrieve them reliably.

Thus, much recent work has focused on decoupling memory storage from model learning and on new ways to access memories. Weston et al. (2014), Sukhbaatar et al. (2015), try to augment the memory of recurrent models, while Lample et al. (2019), Kaur et al. (2021) are recent attempts to integrate and extend the memory of transformer networks efficiently.

We believe understanding the existing knowledge storage and retrieval mechanisms of transformers, will pave the way for better and efficient integration of external memory stores, and models that can store more memories,

more efficiently. Recent works in machine learning interpretability(Voss et al. (2021), Olah et al. (2018)) try to describe the attribution of various features and predictions to various layers and weights for various computer vision models such as ResNet(He et al., 2015). To the best of our knowledge, this is the first work to explore the attribution distributions across different layers and types of knowledge in natural language processing.

3 Types of Knowledge

This paper deals with three types of knowledge, which are explained with the following example.

> "Sarah was visiting [MASK1], the capital [MASK2] France"

Factual knowledge refers to knowledge regarding a certain fact. In the above example, it refers to whether the model can predict the correct word "Paris" at masked position 1, with high probability or not.

Relational Knowledge refers to the knowledge regarding a relation. In the above example, it refers to whether a model is able to identify that it is being asked for the "capital" of *some* nation.

Grammatical Knowledge refers to whether a model knows about the grammatical structure of the language on which it was trained. In the above example it refers to whether the model can predict the correct proposition("of") at masked position 2, with high probability.

This paper concerns itself with the following three questions:

1. How exactly is attribution for Factual knowledge distinct from the attribution of relational knowledge? Which layers are responsible for which knowledge?

2. Do facts in different languages correspond to the same neurons? That is, whether a language model trained on multi-lingual data, has a common representation space for those languages?

3. How is grammatical knowledge dispersed in the various layers of the transformer model? How is this different from relational or factual knowledge?

4 Attribution Strategy for Facts

We prompt the model, using masked sentences, similar in style to the one shown before. Following which, we inspect the logit corresponding to the correct prediction, (i.e., Paris for [MASK1]), and attribute the logit's value to the various neurons in the transformer network using the integrated gradients strategy (Sundararajan et al., 2017) as used in Dai et al. (2021).

This method provides an attribution scheme that is sensitive to changes in input, but invariant to changes in the implementation of the model. At its core, it involves selecting a baseline input(x') and then integrating the gradients of the output($F(x' + \alpha(x - x'))$, with respect to the input, as we increase the input from the baseline(at $\alpha = 0$) to its actual value(at $\alpha = 1$). In our case, the baseline input is chosen as 0, for every neuron.

$$IntegratedGrads_i(x) ::= (x_i - x'_i) \times \int_{\alpha=0}^{1} \frac{\partial F(x' + \alpha(x - x'))}{\partial x_i} d\alpha \quad (1)$$

To find the contribution of a single neuron(i) to the final prediction of, say Paris, we integrate the gradients of the output logit corresponding to Paris, with respect to the activation of that neuron. As we can't compute the integration in Equation 1 exactly, we compute a discrete sum approximation of the integration. After finding the attribution scores of each neuron toward the final prediction this way, we find neurons with maximum or near-maximum scores and the results of Dai et al. (2021) show that these are the ones that store the information leading to the correct prediction.

Furthermore, to remove false positives and identify neurons corresponding to actual facts rather than some spurious signal in the input, Dai et al. (2021) suggest using multiple prompts(corresponding to the same fact) instead of just a single one and only choosing those neurons which have high attribution scores for most of the prompts. These prompts need to have varied structures and grammatical syntax. For example, another prompt corresponding to the (France, Capital, Paris) fact can be:

> France's capital, [MASK] is a hotspot for romantic vacations.

We use the implementation[1] released by EleutherAI, to run our experiments.

[1] https://github.com/EleutherAI/knowledge-neurons

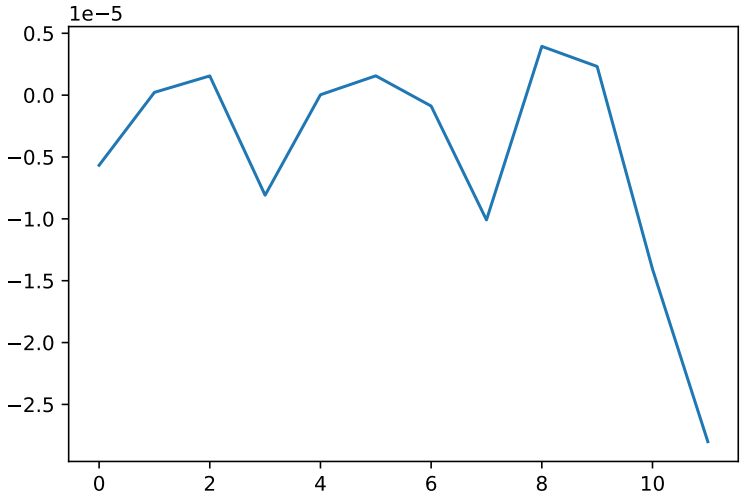


Figure 1: Mean of layer-wise average scores over prompts corresponding to (France, Capital, Paris). Shows the mean attribution of the fact that Paris is capital of France to various layers of the transformer model. The layer number varies across X=axis and mean score along Y-axis.

5 Initial Exploration

Figure 1 shows the average scores over the multiple prompts[2] considered, for each layer of the model. We see that most initial layers consist of approximately 0 attribution scores. Layers 7-8-9 yield high attribution scores, followed by huge(in comparison) negative scores for the final layers. We hypothesize that this behavior is due, mainly to three reasons:

- Initial layers encode low-level syntactic information, which doesn't particularly correspond to the fact of interest.
- The layers 7-8-9 encode higher level information that is of interest to prediction for the fact of our interest, possibly of the relation(viz., "capital") that we are seeking.
- The last few layers, specialize in lots of different directions and only few neurons correspond to the actual fact of interest. The rest of them try

[2]See Appendix A for a complete list of prompts.

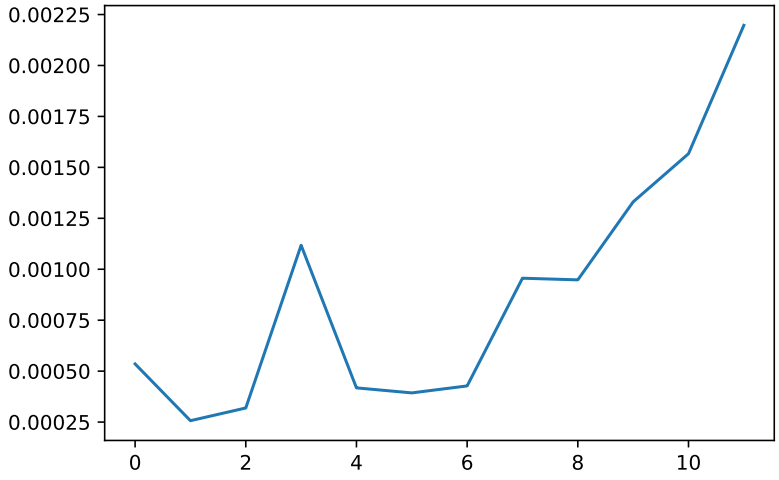


Figure 2: Mean of layer-wise standard deviations of scores over prompts corresponding to (France, Capital, Paris). Layer number on X-axis, standard deviation on Y-axis.

to fit the lower layer representations to the fact they contain(remember, each neuron is a key-value memory) and hence lead to lower attribution score for the actual fact of interest.

The sharp increase in standard deviations from the mean, in the last few layers, in Figure 2 further supports our claim that the last few layers store a variety of different facts.

Figure 3 provides further insight into how knowledge is distributed and retrieved from the transformer. We see that although the mean score attains its maximum around layer 8, the maximum score goes on increasing as we go higher in the model. This suggests that although most of the neurons in higher layers store information highly unrelated to our fact, there does exist some neuron(the one with maximum attribution) that stores exactly the information we need to predict the correct value.

In Figure 4 we see the same kind of behavior for prompts that correspond to a different fact, showing the generalizability of our claims.

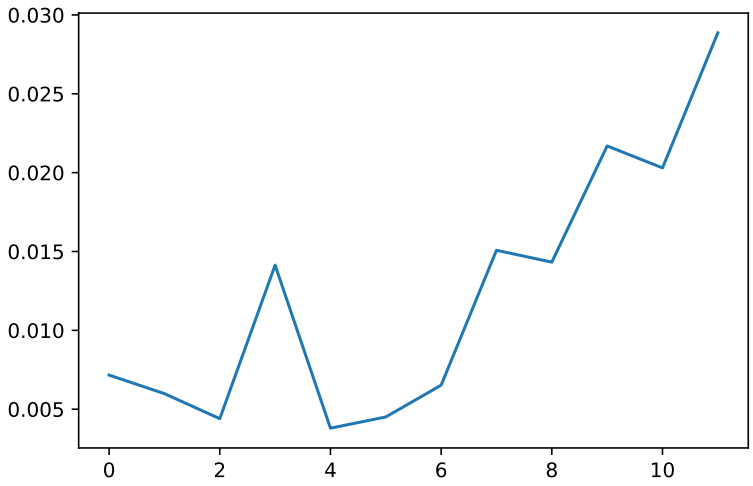
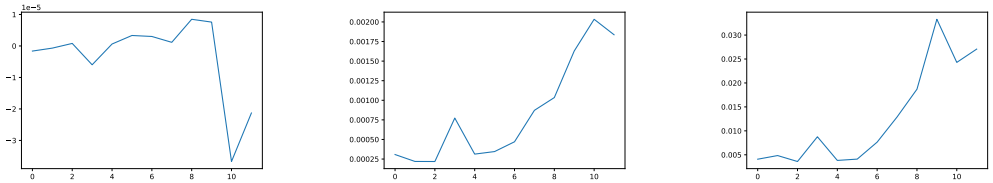


Figure 3: Mean of layer-wise maximum of scores over prompts corresponding to (France, Capital, Paris).Layer number on X-axis, maximum scores on Y-axis.



(a) Average scores (b) Standard deviations (c) Maximum scores

Figure 4: Variations of information content(attribution scores) at various layers for prompts corresponding to the fact (Germany, Capital, Berlin).

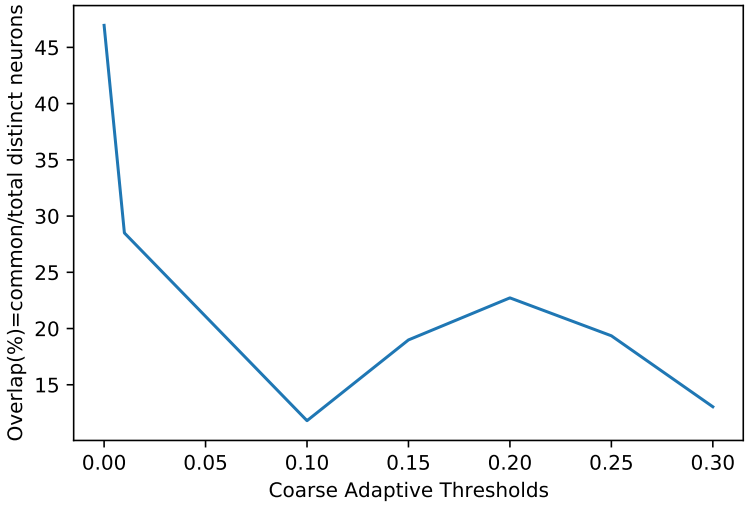


Figure 5: Overlap between knowledge neurons for (France, Capital, Paris) & (Germany, Capital, Berlin) with varying t.

6 Is it the relation or the fact?

Next, we explore whether neurons we find using this attribution method correspond to the fact that "Paris is the capital of France" or the relation between the two objects, i.e., "A is capital of B". Dai et al. (2021) use a threshold t, and only neurons with attribution scores$> t$ are considered the ones corresponding to this fact. In addition, they use another threshold P, and only neurons that have $> t$ attribution scores in more than $P\%$ of the prompts are selected as representing the fact.

We fix P to 50% and vary t to find how the overlap between the neurons identified as representing the fact (France, Capital, Paris) and (Germany, Capital, Berlin) changes. Increasing t for a particular fact has the following effect: at $t = 0$ we are selecting all neurons of the network, with increasing t we start selecting neurons that are more and more particular to the fact. If we continue to increase t further, it is likely that we will lose some of the meaningful neurons as well.

The results are shown in Figure 5. At small thresholds the overlap is quite high(as expected) because we are selecting a large number of neurons for each fact(even unrelated neurons get selected), but the bump in the percentage of

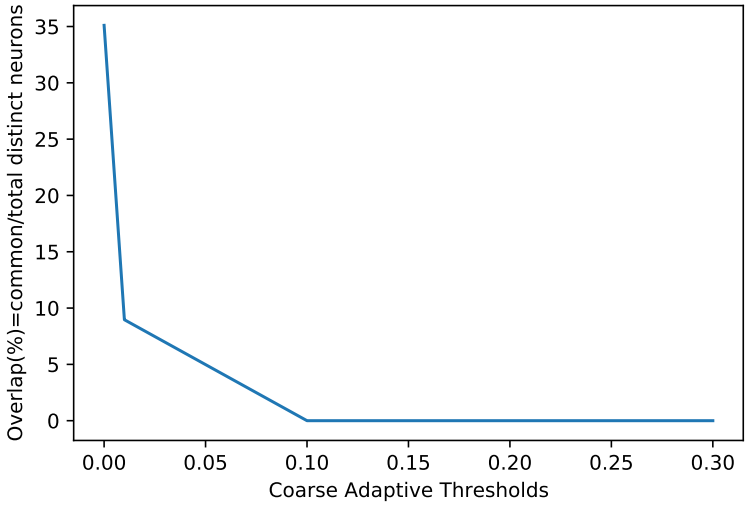


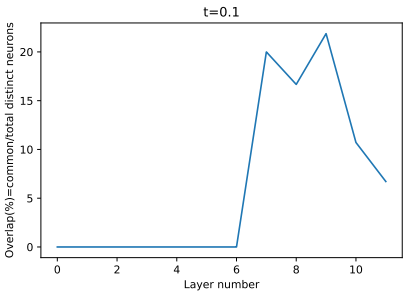
Figure 6: Overlap between knowledge neurons for (France, Capital, Paris) & (Cow, eats, grass) with varying t.

common neurons, **with increased threshold,** t clearly indicates that there are neurons storing the commonality, that is, "A is capital of B", rather than individual information regarding Paris being capital of France or Berlin being capital of Germany.

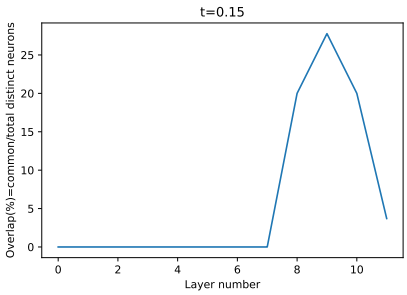
Moreover, predictions at masked positions, for both of these kinds of prompts, are being based on these common "relational" neuron. For comparison, we provide the overlap of commons neurons for two unrelated facts, viz. (France, Capital, Paris) and (Cow, eats, grass), in Figure 6. We see no bump and instead observe a flat line here.

In Figure 7, we see exactly in which layers the common neurons lie. Overall, we observe a trend that:

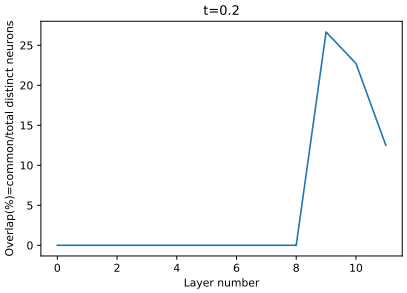
- No model picks any neurons in lower layers
- Most common neurons are found in layer 7-8-9-10. It is likely that the "capital" relation is being realised by the model in these layers.
- In higher layers, we observe a drop in the number of common neurons, again indicating that each fact must belong to different set of neurons



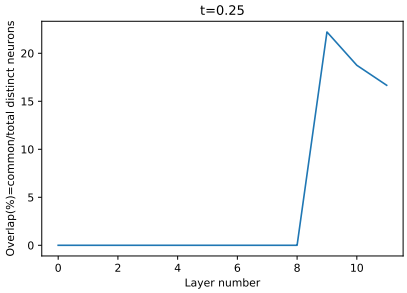
(a) $t = 0.1$



(b) $t = 0.15$



(c) $t = 0.2$



(d) $t = 0.25$

Figure 7: Layer-wise overlap of neurons between the facts (Germany, Capital, Berlin), and (France, Capital, Paris) at different t in the bump in Figure 5.

in higher layer. It is likely that the model is finding the exact capital of the specified country.

7 Do same neurons represent the same fact in different languages?

We use a multilingual transformer encoder model, **m-BERT** (Devlin et al., 2019) to check whether facts in different languages access information from the same neurons? We construct two sets of prompts corresponding to the relational-fact (France, Capital, Paris) in English, and translate one set of prompts to French to construct a set of prompts corresponding to the same fact, but in a different language.

We calculate the overlap between the neurons used for the two sets of English prompts and plot it in blue. We also measure the overlap between the neurons selected for the French prompts and those for the combined set of English prompts and plot this overlap in orange, in Figure 8. We observe that both overlaps show similar trends and are close to each other. In general, as expected, they show a falling trend as we increase t. This closeness of the two curves in Figure 8 is an attestation to the ability of transformers to efficiently store multilingual information, by taking information in different languages to a common representation space (indicated by the common neurons). This provides further force to the claim of Del and Fishel (2021), that different languages do share a common representation space.

The layer-wise analysis is shown in Figure 9. For $t \in [0.1, 0.25]$ we observe the same trend as before(Figure 5), where overlap increases first, then decreases in the last few layers. A difference we observe is that the overlap stays non-zero for much higher t's than the ones we obtained while comparing measuring overlap for (France, Capital, Paris) and (Germany, Capital, Berlin). This indicates that the same facts in different languages share many more common neurons of **m-BERT**, than those shared between facts representing the same relation.

For $t > 0.25$, the trend is different and we see no drop in higher layers, probably because we only choose neurons with very high attribution scores, which means that they may already correspond to very specialized information. At around 0.5, we see that no overlap exists between the neurons used for different languages.

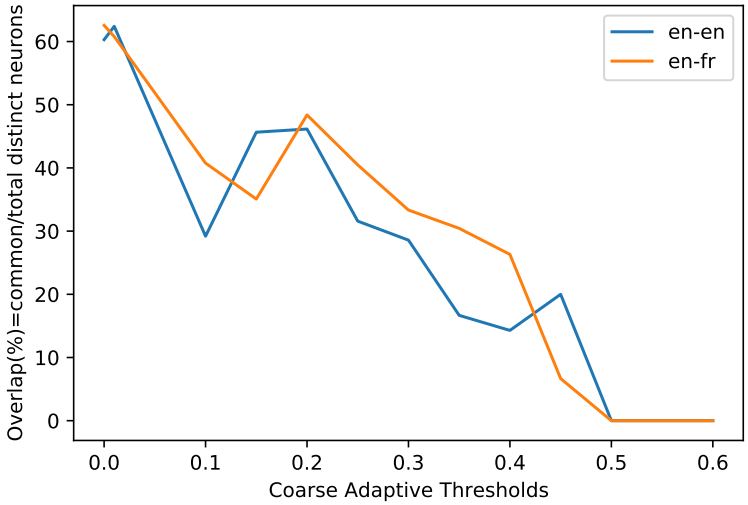


Figure 8: Variation of overlap of attributed neurons for facts in different languages and same languages with varying t.

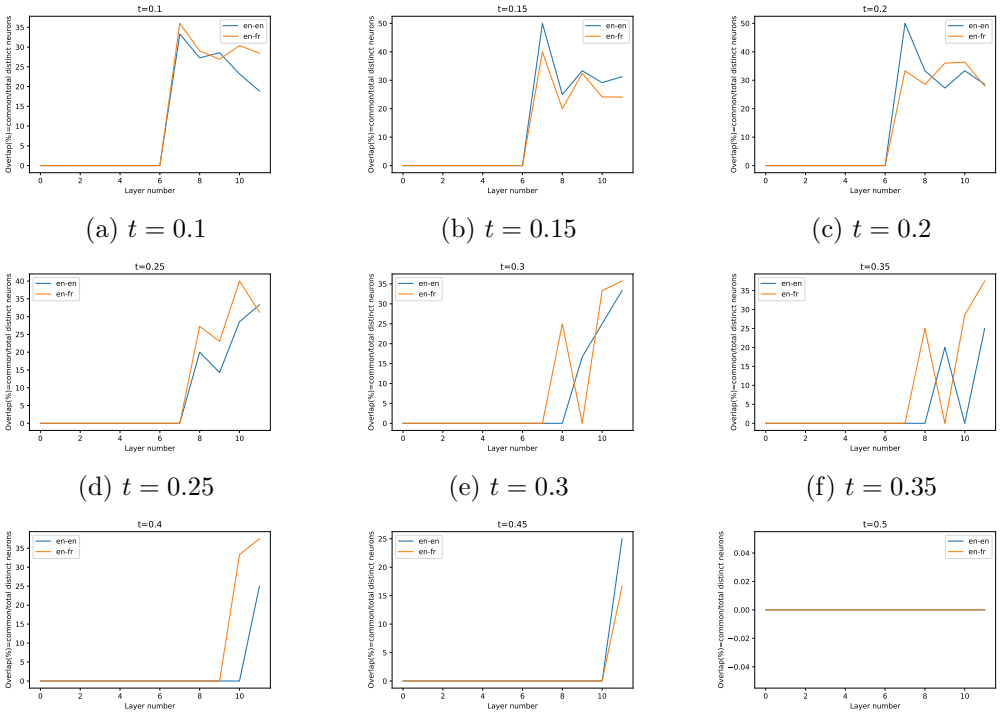
8 Grammatical Knowledge Attribution

As we have prompts checking for factual knowledge attribution, we can also design prompts for grammatical knowledge attribution. We investigate the properties, of grammatical knowledge attribution next. We use the Number Agreement dataset curated by Gulordava et al. (2018). We use these data with BERT, in a way similar to what Goldberg (2019) do in their repository[3].

As this dataset consists mainly of non-sensical sentences and lacks semantic or lexical cues, we can be sure that we are indeed considering attribution of grammatical knowledge. Moreover, as there are correct(**good**) labels, as well as incorrect(**bad**) ones(corresponding to the wrong number of subject/verb) provided, we can check which neurons are responsible for the correct prediction, and which push the prediction in the wrong direction. We also show variation of results as the number of attractors(words that distract the model from predicting the correct number) present in the sample prompt, change.

Figure 10 shows the layer-wise variation of attribution scores for predicting the word with the correct number, and predicting the word with incorrect

[3] https://github.com/yoavg/bert-syntax

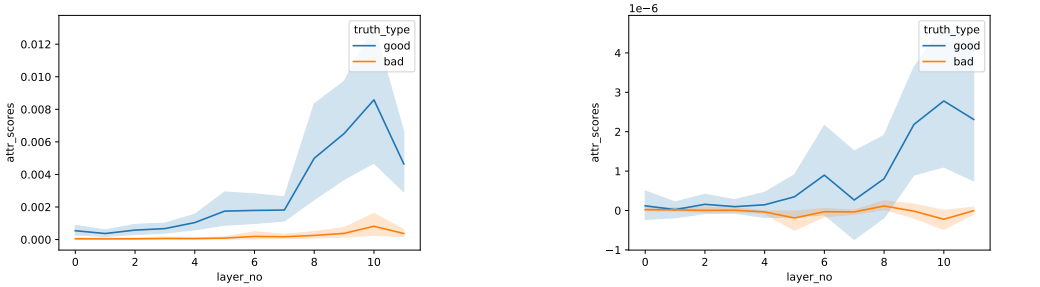


(a) $t = 0.1$ (b) $t = 0.15$ (c) $t = 0.2$

(d) $t = 0.25$ (e) $t = 0.3$ (f) $t = 0.35$

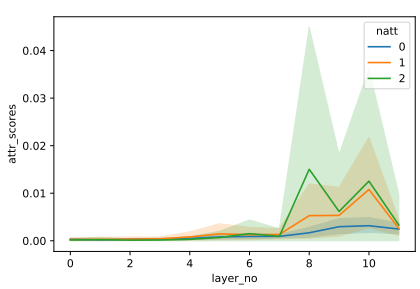
(g) $t = 0.4$ (h) $t = 0.45$ (i) $t = 0.50$

Figure 9: Layer-wise overlap of neurons between facts in same language and different language at different t.

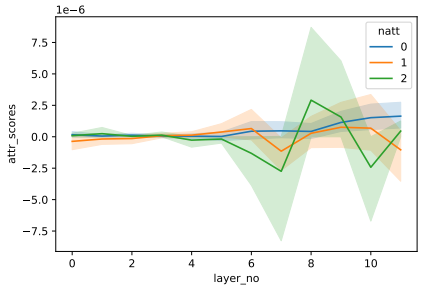


(a) Max. attribution scores in each layer (b) Mean attribution scores in each layer

Figure 10: Variation of attribution scores for predicting correct number(in blue) and incorrect number(in orange). The scores are averaged over all examples in the dataset and shaded regions indicate confidence intervals.



(a) Max. attribution scores in each layer



(b) Mean attribution scores in each layer

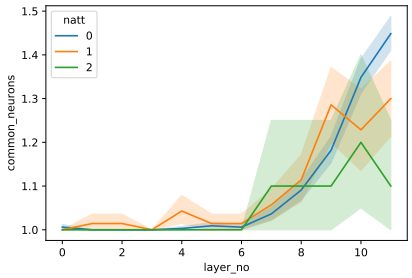
Figure 11: Variation of attribution scores for examples with different number of attractors. The scores are averaged over all examples in the dataset and shaded regions indicate confidence intervals.

number, at the masked position.

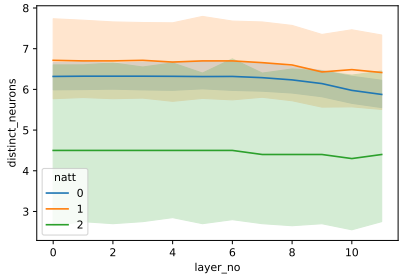
First, we observe that the attribution scores for correct predictions are almost always higher than the attribution scores for incorrect predictions, showing the ability of the model to discriminate between the two, beginning as early as the 4^{th} layer. Second, we see that the highest attribution scores in the case of grammatical knowledge are around 0.1, compared to around 0.3 in the case of factual knowledge(Figure 3 and Figure 4c). This probably is an indication that grammatical knowledge is more dispersed in the model than factual knowledge.

Figure 11 shows the variation of the attribution scores for examples categorised according to various number of attractors present. We observe that the variation of attribution scores increases significantly as we increase the number of attractors, showing that the model is indeed getting confused by the presence of attractors. In addition, the variation increases as we go to higher layers.

Next, we find refined neurons for the correct and incorrect prediction of each sample in the dataset, using an adaptive threshold of 0.5. We calculate the number of neurons common for the correct and incorrect predictions(Figure 12a) and number of neurons that store information for either the correct answer or the incorrect answer("decided" neurons). The common neurons("undecided" neurons) indicate those neurons that do not differentiate between the correct and incorrect predictions, while the distinct ones are the ones which only support only one of correct or the incorrect answer.



(a) Common refined neurons for correct and incorrect prediction of the same sample



(b) Distinct refined neurons for correct and incorrect prediction of the same sample

Figure 12: Variation of number of refined neurons for examples with different number of attractors. The numbers are averaged over all examples in the dataset and shaded regions indicate confidence intervals.

We see that these "undecided" neurons occur more towards the later layers(Figure 12a) but are much less in number, than the "decided" neurons shown in Figure 12b.

9 Conclusion and Future Work

We presented various interesting patterns on the location of knowledge of various kinds in various parts of transformer networks and the interactions of various kinds and languages in which knowledge can occur using the Knowledge Neurons(Dai et al., 2021) attribution technique.

We leave it for future work to explore other attribution techniques(Ribeiro et al. (2016), Lundberg and Lee (2017), Algebraist (2020)) and on other language models like GPT(Brown et al., 2020), which have been trained with a different objective(CLM) than the MLM objective of BERT model we have used.

References

Nost Algebraist. Interpreting gpt: The logit lens, 31 August, 2020. URL https://www.lesswrong.com/posts/AcKRB8wDpdaN6v6ru/interpreting-gpt-the-logit-lens.

Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. Language models are few-shot learners, 2020. URL https://arxiv.org/abs/2005.14165.

Damai Dai, Li Dong, Yaru Hao, Zhifang Sui, and Furu Wei. Knowledge neurons in pretrained transformers, 2021.

Maksym Del and Mark Fishel. Establishing interlingua in multilingual language models, 2021. URL https://arxiv.org/abs/2109.01207.

Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186. Association for Computational Linguistics, June 2019. doi: 10.18653/v1/N19-1423. URL https://aclanthology.org/N19-1423.

Mor Geva, Roei Schuster, Jonathan Berant, and Omer Levy. Transformer feed-forward layers are key-value memories, 2020. URL https://arxiv.org/abs/2012.14913.

Yoav Goldberg. Assessing bert's syntactic abilities. *arXiv preprint arXiv:1901.05287*, 2019.

Kristina Gulordava, Piotr Bojanowski, Edouard Grave, Tal Linzen, and Marco Baroni. Colorless green recurrent networks dream hierarchically. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1195–1205, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1108. URL https://aclanthology.org/N18-1108.

Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015. URL https://arxiv.org/abs/1512.03385.

Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 9(8):1735–1780, 1997.

J J Hopfield. Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Sciences*, 79(8):2554–2558, 1982. doi: 10.1073/pnas.79.8.2554. URL https://www.pnas.org/doi/abs/10.1073/pnas.79.8.2554.

Jivat Neet Kaur, Sumit Bhatia, Milan Aggarwal, Rachit Bansal, and Balaji Krishnamurthy. No need to know everything! efficiently augmenting language models with external knowledge. In *Workshop on Commonsense Reasoning and Knowledge Bases*, 2021.

Guillaume Lample, Alexandre Sablayrolles, Marc'Aurelio Ranzato, Ludovic Denoyer, and Hervé Jégou. Large memory layers with product keys, 2019.

Scott Lundberg and Su-In Lee. A unified approach to interpreting model predictions, 2017. URL https://arxiv.org/abs/1705.07874.

Chris Olah, Arvind Satyanarayan, Ian Johnson, Shan Carter, Ludwig Schubert, Katherine Ye, and Alexander Mordvintsev. The building blocks of interpretability. *Distill*, 2018. doi: 10.23915/distill.00010. https://distill.pub/2018/building-blocks.

Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. "why should i trust you?": Explaining the predictions of any classifier, 2016. URL https://arxiv.org/abs/1602.04938.

Sainbayar Sukhbaatar, Arthur Szlam, Jason Weston, and Rob Fergus. End-to-end memory networks, 2015. URL https://arxiv.org/abs/1503.08895.

Mukund Sundararajan, Ankur Taly, and Qiqi Yan. Axiomatic attribution for deep networks, 2017.

Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Lukasz Kaiser, and Illia Polosukhin. Attention is all you need, 2017. URL https://arxiv.org/abs/1706.03762.

Chelsea Voss, Nick Cammarata, Gabriel Goh, Michael Petrov, Ludwig Schubert, Ben Egan, Swee Kiat Lim, and Chris Olah. Visualizing weights. *Distill*, 2021. doi: 10.23915/distill.00024.007. https://distill.pub/2020/circuits/visualizing-weights.

Jason Weston, Sumit Chopra, and Antoine Bordes. Memory networks, 2014. URL https://arxiv.org/abs/1410.3916.

Appendices

A Prompts used

A.1 English Prompts-(France, Capital, Paris)-Set 1

1. "Sarah was visiting [MASK], the capital of france",
2. "The capital of france is [MASK]",
3. "[MASK] is the capital of france",
4. "France's capital [MASK] is a hotspot for romantic vacations",
5. "The eiffel tower is situated in [MASK]",
6. "[MASK] is the most populous city in france",
7. "[MASK], france's capital, is one of the most popular tourist destinations in the world"

A.2 English Prompts- (France, Capital, Paris)- Set 2

1. "I have always wanted to know what the capital of France is... and now I know it is [MASK] !",
2. "Do you even know the name of the capital of France? It's not Berlin, it's [MASK].",
3. "The city of [MASK] was made the capital of France in 987 A.D.",

4. "Claude Monet, the famous painter was also born in [MASK], the now capital of France.",

5. "As is London the gem of Britain, so is [MASK], the centre of France.",

6. "At the beginning of 20th century, [MASK], being the capital of France, was the largest catholic city in the world.",

7. "The city of [MASK], housing the largest business district in Europe, La Défense, is the rightful of its place as the capital of France.",

A.3 French Prompts-(France, Capitale, Paris)

1. "Sarah était en visite à [MASK], la capitale de la France",

2. "La capitale de la france est [MASK]",

3. "[MASK] est la capitale de la France",

4. "[MASK], la capitale de la France, est un haut lieu des vacances romantiques",

5. "La tour eiffel est située à [MASK]",

6. "[MASK] est la ville la plus peuplée de France",

7. "La capitale de la France, [MASK], est l'une des destinations touristiques les plus populaires au monde",

A.4 English Prompts-(Germany, Capital, Berlin)

1. "Why would Sarah ever visit [MASK]? Doesn't she know that its the capital of Germany!?",

2. "Hitler lived most of its life in [MASK], the now capital of Germany.",

3. "[MASK] is the capital of Germany",

4. "Germany's capital [MASK] is a hotspot for tourists",

5. "The Brandenburg Gate is situated in [MASK]",

6. "[MASK] is the most populous city in Germany",

7. "[MASK], Germany's capital, is visited by a lot of Dutch people every year",

A.5 English Prompts-(Cow, Eats, Grass)

1. "Yesterday, I saw a cow eating [MASK]",

2. "Cows love to eat [MASK]",

3. "Herbivores like cows, that eat [MASK], usually have a separate stomach to ruminate",

4. "All the nutrients in cow's milk comes from the [MASK] it eats",

5. "Some people were found to be living worse than cows: eating breads made out of [MASK]",

6. "The cows in praries look most beautiful when gnawing on [MASK], just before the sunset",

7. "Spherical cows eat get hungry too! They too eat [MASK]."

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

FLOATING SOLAR PHOTOVOLTAIC SYSTEM

Ayush Goel[*1], Deepanshu Yadav[*2], J.P. Kesari[*3]

[*1,2]Student, Department Of Mechanuical Engineering Delhi Technological University (Formerly Delhi College Of Engineering) Bawana Road, Delhi – 110042, India.

[*3]Guide, Department Of Mechanuical Engineering Delhi Technological University (Formerly Delhi College Of Engineering) Bawana Road, Delhi – 110042, India.

ABSTRACT

India has done a notable job in terms of deployment of renewable energy-based installations, increasing to almost 3.5 folds in the last 5–6 years, with most of the capacity coming from onshore wind and solar photovoltaic (PV) based installations. India's grid-connected solar PV sector currently is majorly dominated by the ground-based installations (93%1) while the balance is contributed by rooftop based solar PV installations. The cost of installation of utility-scale solar PV in the country has declined by 84% between 2010-2018, making India world's topmost country to achieve the lowest installation cost for utility- scale solar PV. It is well known that solar PV deployment is very land intensive and scaling up the project sizes requires large portion of contiguous land parcels, which becomes challenging in many situations. In order to keep pace of development commensurate with the country's targets for solar capacity additions, alternatives are required to be established and explored. Floating solar PV (FSPV) or floatovoltaics is one such substitute, which has started getting attention globally and is expected to grow strongly over the coming years.

It is assessed that the yearly capacity addition may rise from the current installed of 1.314 GWp in 2018 to 4.6 GWp by 2022. At present, China is the leading international market followed by Japan and South Korea. India also has very bright outlooks to develop FSPV projects due to availability of large water bodies.

As a technology, Floating Solar PV is in a very nascent stage of development in India. Until now, only a few projects with cumulative capacity of 2.7 MW have been installed. However, over 1.7 GW capacity projects are reported to be in various processes of development. FSPV market appears to crawl forward to make its presence felt in India and the taxes and tariffs discovered through bids have already shown rapid reductions. So far large-to-medium size man-made inland waterbodies seems to have attracted initial interest to install FSPV based power plants, but all these waterbodies were created to serve various purposes like – irrigation, water supply, fishing, hydroelectric, navigation, etc., and this warrants great deal of diligence to balance out various usages of these waterbodies on the basis of accurate information.

Keywords: FSPV, Floating Solar PV, Solar Panel, Floatovoltaics.

I. INTRODUCTION

Solar power is a smooth and renewable shape of power compared to fossil fuels. With the sturdy commitments of the country wide governments round the arena closer to greenhouse gases (GHG) reduction, sun Photovoltaics (PV) is a smooth generation technique to lessen the Greenhouse Gas emissions withinside the electricity quarter. In many nations, land recourses are constrained for large scale ground-set up sun PV structures. Moreover, rooftop regions in housing, business and commercial homes won't be vital for rooftop sun solution. In this context, floating sun PV structures are the correct ecological opportunity solutions. Floating sun photovoltaics is coined as "floatovoltaics".

As in step with the International Energy Agency, the deployment of sun photovoltaic had been globally at top in 2020 withinside the attention of 90% develop in call for of renewable electricity. Photovoltaic (PV) are capable of at once convert daylight to electricity. From photovoltaic factor of view, there are 4 predominant elements influencing the output PV power yield. First comes the supply of power that is the daylight, then is the converter that is the PV cell, furthermore the quantity of time for which the Photovoltaic unit can function, and eventually the dimensions of place that this generation is auxiliary on. We do now no longer have plenty manage on daylight. In addition to that, PV converter performance is touching its most theoretical performance. And this is why researchers have now placed greater attempt into investigating tactics to reinforce the lifetime of PV and additionally searching into opportunities to add PV on or included it into any feasible floor. Since 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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

performance of PV modules is low, they take big quantity of place, which may be applied for different vital requirements of human kind, consisting of meals and accommodation. The call for for meals, accommodation, and inexperienced power is growing with growth in international populace therefore, there may be a opportunity that meals and power sectors may compete or already are competing over place. This necessarily attracts the eye to some other in large part to be had floor place, the water. Simply putting any kind of PV device on pinnacle of water bodies, consisting of lakes, reservoirs, hydroelectric dams, commercial and irrigation ponds, and coastal lagoons, is referred to as floating PV (FPV) or floatovoltaics.

1.1 WORLD ENERGY SCENARIO

1.1.1 UNITED STATES OF AMERICA

The Solar Photovoltaics Supply Chain Review explores the worldwide sun photovoltaics (PV) deliver chain and possibilities for growing U.S. production ability. The evaluation concludes that, with tremendous economic help and incentives from the U.S. authorities in addition to strategic moves centered on workforce, production, human rights, and trade, America should re-set up a sturdy home sun production deliver chain and grow to be a aggressive chief in a worldwide sun industry. This should result in remarkable blessings for the weather in addition to for U.S. workers, employers, and the economy. Produced via way of means of the U.S. Department of Energy (DOE) Solar Energy Technologies Office (SETO) with help from the National Renewable Energy Laboratory (NREL) and launched on February 24, 2022, that is one in all a sequence of power quarter commercial reviews directed thru President Biden's Executive Order 14017 "America's Supply Chains." The Executive Order will assist the U.S. federal authorities to construct greater steady and various U.S. power deliver chains – facilitating more home production, an acceleration in smooth power, a variety of deliver, integrated redundancies, ok stockpiles, secure and steady virtual networks, and a international-magnificence American production base and workforce.

1.1.2 CHINA

China is a leading country in terms of highest installation capacity accounting for more than 960 MWp out of global total of 1314 MWp. The country is also home to the world's largest FSPV plant of 150 MW capacity installed at Huainan, south Anhui province. Most of the FSPV plants in China are installed at unused mining ponds.

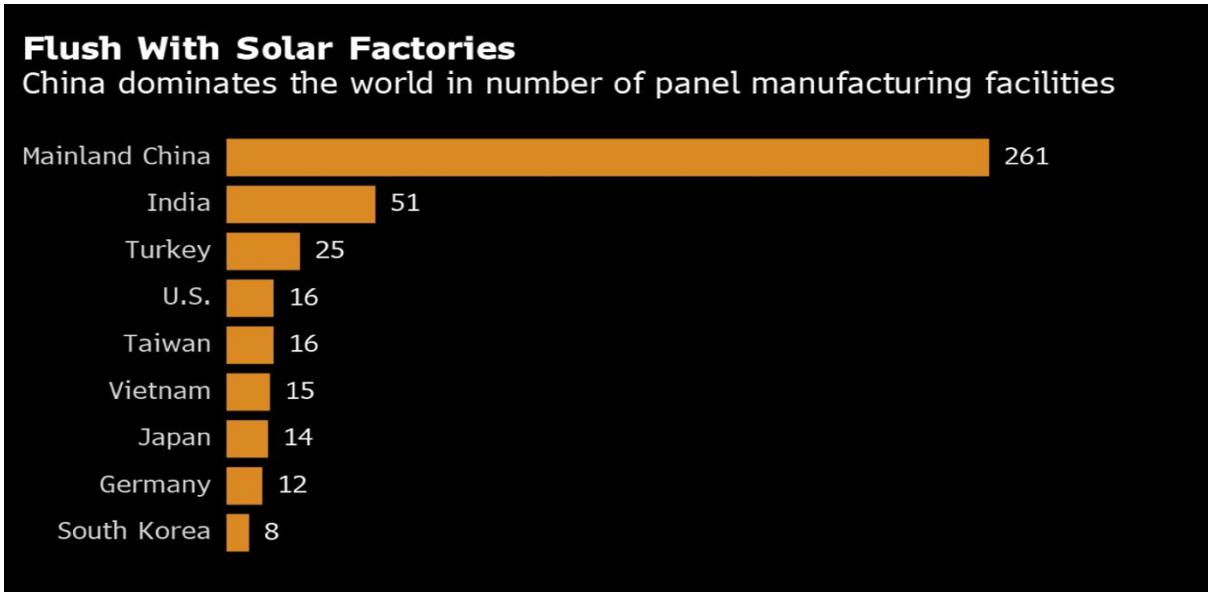


Figure 1: Solar Panel manufacturing facilities

1.1.3 INTERNATIONAL SOLAR ALLIANCE (ISA)

The release of the International Solar Alliance (ISA) became introduced via way of means of H.E. Mr. Narendra Modi, the Hon'ble Prime Minister of India and H.E. Mr. Francois Hollande, former Hon'ble President of France on thirtieth November 2015, on the twenty first consultation of United Nations Climate Change Conference of the Parties (COP-21) in Paris, France. Former UN Secretary-General Ban Ki-moon attended the release, along

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

the Heads of approximately a hundred and twenty countries who affirmed their participation withinside the Alliance to commit efforts for advertising of sun power. The ISA became conceived as a joint attempt via way of means of India and France to mobilize efforts towards weather alternate thru deployment of sun power solutions. It became conceptualized at the sidelines of the twenty first Conference of Parties (COP21) to the United Nations Framework Convention on Climate Change (UNFCCC) held in Paris in the year 2015. With the change of its Framework Agreement in 2020, all member states of the United Nations are actually eligible to enroll in the ISA. At present, a hundred and one nations are signatories to the ISA Framework Agreement, of which eighty nations have submitted the vital contraptions of ratification to grow to be complete individuals of the ISA.

1.2 INDIAN SCENARIO

The FSPV as a technology is still in the nascent stages of development in India. The journey started with a 10kW FSPV plant on a pond in Rajarhat, Kolkata in 2015. The project was part of a research activity sponsored by the Ministry of New and Renewable Energy (MNRE). In 2016, NTPC installed country's largest 100kW plant on a reservoir of its combined cycle power plant situated in Kerala's Kayamkulam district. Later in December same year, Kerala State Electricity Board started its operation of 500kW plant at Banasura Sagar reservoir in Wayanad district replacing NTPC's 100kW as a largest FSPV-based plant. The plant is actually a scaled-up version of the 10kW plant commissioned in January 2016 at the same location. The plant was able to bring some confidence to FSPV promoters by successfully surviving the recent flood ravage in the state. The recently commissioned 2MW project at Visakhapatnam, Andhra Pradesh has now the largest FSPV-based plant commissioned in the country till date and with this the total installed capacity of FSPV has becomes 2.7 MW. The FSPV sector is now getting a lot of attention in the country, which can be seen via an increase in the numbers of tenders that are released in the past 2 years. At the moment there is more than 1700 MW worth of projects, which are in various stages of development and more are in pipeline making the outlook very positive for this new segment.

II. REVIEW OF LITERATURE

2.1 Analytical Framework and Related Literature

Integrated Renewable energy systems and renewable energy systems are used to take benefit of available renewable supply (Vishnupriyan & Manoharan, 2018). Renewable energy sources play a crucial role in the production of electrical energy (García et al., 2014), because they function to overpower dependence on non-renewable resources (Sahu & Raheman, 2020). Renewable energy sources (RE) are the best alternative available to fulfil the demand for electricity. The main sources of Renewable Energy based systems are solar Photovoltaic systems, wind power generation systems, fuel cells (FC), micro-turbines etc. Renewable Energy sources are gaining more and more attention for industrial as well as domestic applications due to their advantageous features such as reliability, availability and environmental friendliness. Solar energy is deliberated as a promising, profitable and reliable energy source. It has countless advantages such as pollution free, long life, low maintenance etc. (Gupta et al., 2016). Solar energy is the most abundant source of energy that can fulfil community requirements that come from sustainable economic development (Husain et al., 2018). The vogue of growing rapid PV energy use is associated with increased efficiency of solar cells as well as improvements in manufacturing of solar panel technology (Dondi et al., 2008). In most scenarios, desalination plants powered by non-renewable energy sources count on fossil fuel energy to power solar generators, resulting in greenhouse gas emissions and environmental pollution (Elmaadawy et al., 2020). Therefore, real power control is also an efficient way to synchorize voltage interactions in a network distribution (Safayet et al., 2017) (Rottondi et al., 2016). Also management strategies (Merei et al., 2016), for electricity efficiency (Monjezi et al., 2020), the price of installing a solar tracking system, various moving parts, motorization and maintenance costs must all accounted while considering an increase in production of power (Hasan & Dincer, 2020).

2.2 Relevance with Production Engineering

Being energy self-sufficient is a crucial national objective. In case of solar energy, this can be achieved sans sacrificing on competitiveness through appropriate strategic policies and global vision to assist smart production and creation of effective supply chain. With growing attention on green and cleaner sources of

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

energy in the country, solar photovoltaic (PV) manufacturing in India is getting a stimulus with laid-back acceptance across potential users and these further leverages specific integral advantages including low cost of HR capital and prevalent penetration of smart production programs throughout industry which relaxes capital investment. Solar Energy will have a vital role to play in fulfilling India's energy security needs in the coming years. The increasing energy demands of India and the focus on clean energy has generated unique opportunities for the solar energy sector in India. India presents a massive marketplace for the growth and penetration of solar energy; Eliminates risk associated with installations which have a high operating life since the main suppliers being indigenous enforcement of obligations contractually and legally becomes more efficient and effective vis-a-vis global suppliers and specially from project lenders and investing entities point of view prevents undermining of highly potential sunshine sector in India.

III. TECHNOLOGY ANALYSIS

3.1 COMPONENTS OF FLOATING SOLAR PV SYSTEM

3.1.1 Solar Module / PV System

The essential part of the floating sun photovoltaic device is sun PV modules and much like traditional sun tasks generally, poly or monocrystalline or skinny movie sun panels are used for the set-up of the project. A single module can produce best a restrained quantity of energy; maximum installations include a couple of modules. A photovoltaic device normally consists of a panel or an array of sun modules, a sun inverter, and now and again a battery and a sun tracker and inter-connection wiring.

3.1.2 Pontoon / Floating Structure

A pontoon is defined as a flotation structure which has enough buoyancy to float on water and support a heavy load. The structure is designed such that it can hold a number of panels. It allows the installation of the photovoltaic module. This is the most crucial component of floating solar photovoltaic system which supports all the necessary components. Therefore, it is very important to select the appropriate materials for the floating structure.

3.1.3 Mooring system:

A mooring refers to an everlasting shape to which a floating shape may be secured. A floating shape is secured to a mooring to stop unfastened motion of the floating shape at the water. An anchor mooring fixes a floating shape's role relative to some extent on the lowest of a waterway, without connecting the floating shape to shore. Since FSPV flowers are established on water bodies, any versions in water degrees caused with the aid of using monsoon, wind pace or increase/lower in water amount can be intricate for the flowers. To keep away from this situation, FSPV flowers are anchored via mooring systems. It has the cap potential to modify to water stage fluctuations whilst retaining its role with inside the southward direction. The placement of a mooring device ought to do not forget the location, bathymetric, soil conditions, and water-stage versions.

3.1.4 Underwater Cable

It is responsible for the transfer of the generated energy; therefore, cable routing and its management requires cautious planning. Due to their outdoor usage, solar cables are designed uniquely to be resistant against UV radiation and extremely high temperature fluctuations and are generally unaffected by the weather. The cables are routed by two different ways - either via floating on water surfaces or via submarine cables.

3.1.5 INVERTERS

Similar to a traditional solar plant, DC power generated from solar PV modules is taken to the inverter through a series of combiner boxes and finally converted into AC power. A developer may select multiple string inverters or central inverters. Depending upon scale and distance from shore, inverters can be placed either on a separate floating platform or on land.

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

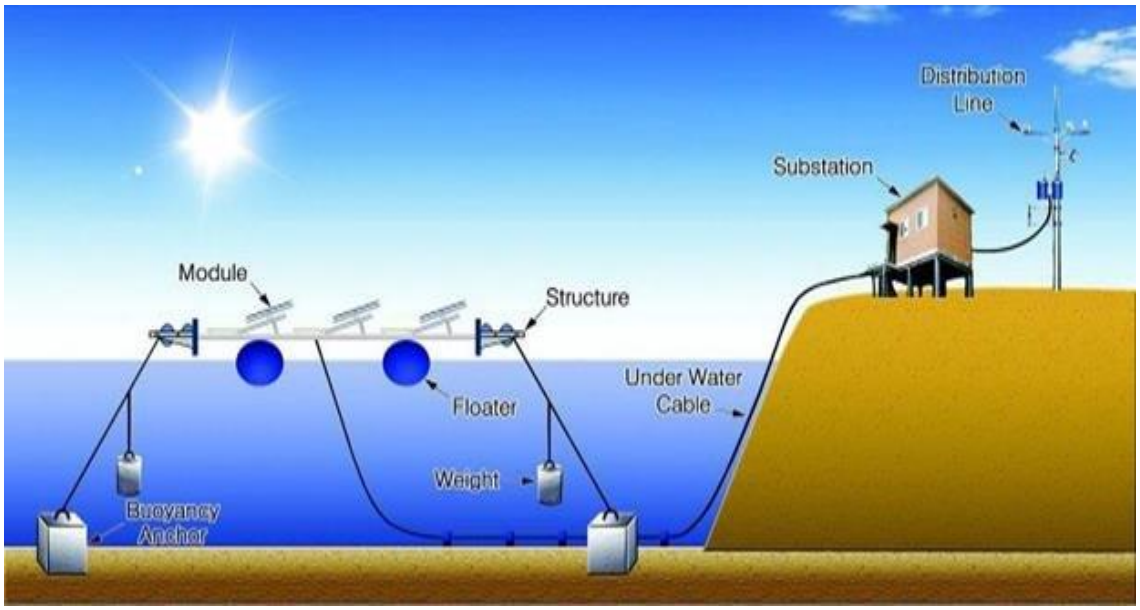


Figure 2: Layout of a typical floating PV system

3.2 TECHNOLOGY ADVANTAGES

The most important parameter considered for the performance evaluation of the floating solar PV is the module effective conversion efficiency in operative conditions, which affects the electricity generation, therefore, the most valuable product of the component. The conversion efficiency of a PV module is given by the ratio between the generated electrical power and the incident solar radiation intensity. The expression is as follows:

$$\eta_{el} = \frac{P_{max}}{S \times A_{PV}} \times 100$$

Where,

η_{el} is the efficiency of electricity produce (%),

P_{max} is the maximum power produce by PV system (W),

S is the solar radiation strength fall on the PV module (W/m^2)

A_{pv} is the area of PV module on that solar radiance fall on the surface (m^2)

3.3 TECHNOLOGICAL DISADVANTAGES

Few of the several disadvantages of the floating photovoltaic system that need to be taken into consideration include the following:

- The floating PV system is more exposure to hydraulic and weather conditions, then may result in unstable power output.
- Fishing and transportation activities may be affected by the floating system.
- Located within the water environment could lead to the corrosion of modules and structures, thus could reduce the system lifetime.

3.4 Technical Design Parameters

- **Energy demand**

First, calculate the total energy demand per day (kWh/day). This calculation is done by multiplying each electrical equipment power by time required to operate and number of items.

$$\mathbf{E = P \times n \times t\ (kWh/day)}$$

Where,

E is energy demand,

P is power in watt (W) of each electrical device,

n is No. of items,

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

t is time required to operate in hours (h).

- **Weather forecasting**

After energy calculation, the weather data like monthly and daily average insulation, minimum and maximum daily average temperature, monthly or daily average wind speed, has been observed from the NASA website or other solar simulation software by implanting the latitude and longitude of the installation site of about 10 to 20 years' span.

- **Cell temperature of Floating PV system from weather data**

The sea temperature, sea wind speed, PV cell temperature on land, PV cell temperature in the sea were determined by the following equations.

$$T_w = 5.0 + 0.75T_a$$

Where T_w = sea temperature (°C) and Ta = air temperature (°C).

$$V_{wsea} = 1.62 + 1.17 \times V_{wland}$$

Where V_{wsea} is sea wind speed, V_{wland} is land wind speed m/s.

$$T_c = 0.943 \times T_a + 0.0195 \times G - 1.528 \times V_{wland} + 0.3529$$

Where T_c is PV cell temperature on land G is global solar radiation at STP which 1000 W/m2.

$$T_{cw} = 0.943 \times T_w + 0.0195 * G - 1.528 \times V_{wsea} + 0.3529$$

Where T_{cw} is PV cell temperature in the sea.

- **De-rated Daily Output Energy of Offshore Floating PV Array**

$$E_{PV} = W_p\,(f_{dc/ac})(G/G_{STC})[1 + \beta(T_c - T_{STC})]$$

Where,

E_{PV} is energy output from PV,

W_p is power output at STP,

$f_{dc/ac}$ is DC to AC de-rating factor,

G is incident solar radiation,

G_{STC} is solar radiation at STP,

β is power temperature coefficient %/°C.

- **Size of PV Array**

$$W_p = \frac{E_L}{\left(\frac{G}{G_{STC}}\right)(f_{dc/ac})(f_{temp})}$$

Where, $f_{temp} = 1+\beta(T_C - T_{STC}) = 1+\beta(T_{Cw} - T_{STC})$.

No. of modules in parallel N_{mp}

$$N_{mp} = \frac{W_p}{\text{peak power of each module}}$$

No. of modules in series N_{ms}

$$N_{ms} = \frac{\text{System DC voltage}}{\text{module voltage}}$$

- **Size of Inverter**

In the first step the size of the Inverter is determine by the real power drawn from all appliance run at a same time. In second to start the heavy motors must consider the inrush current and compensate the power by multiplying the power by a factor by 3 otherwise we use 1.25 standard safety factor.

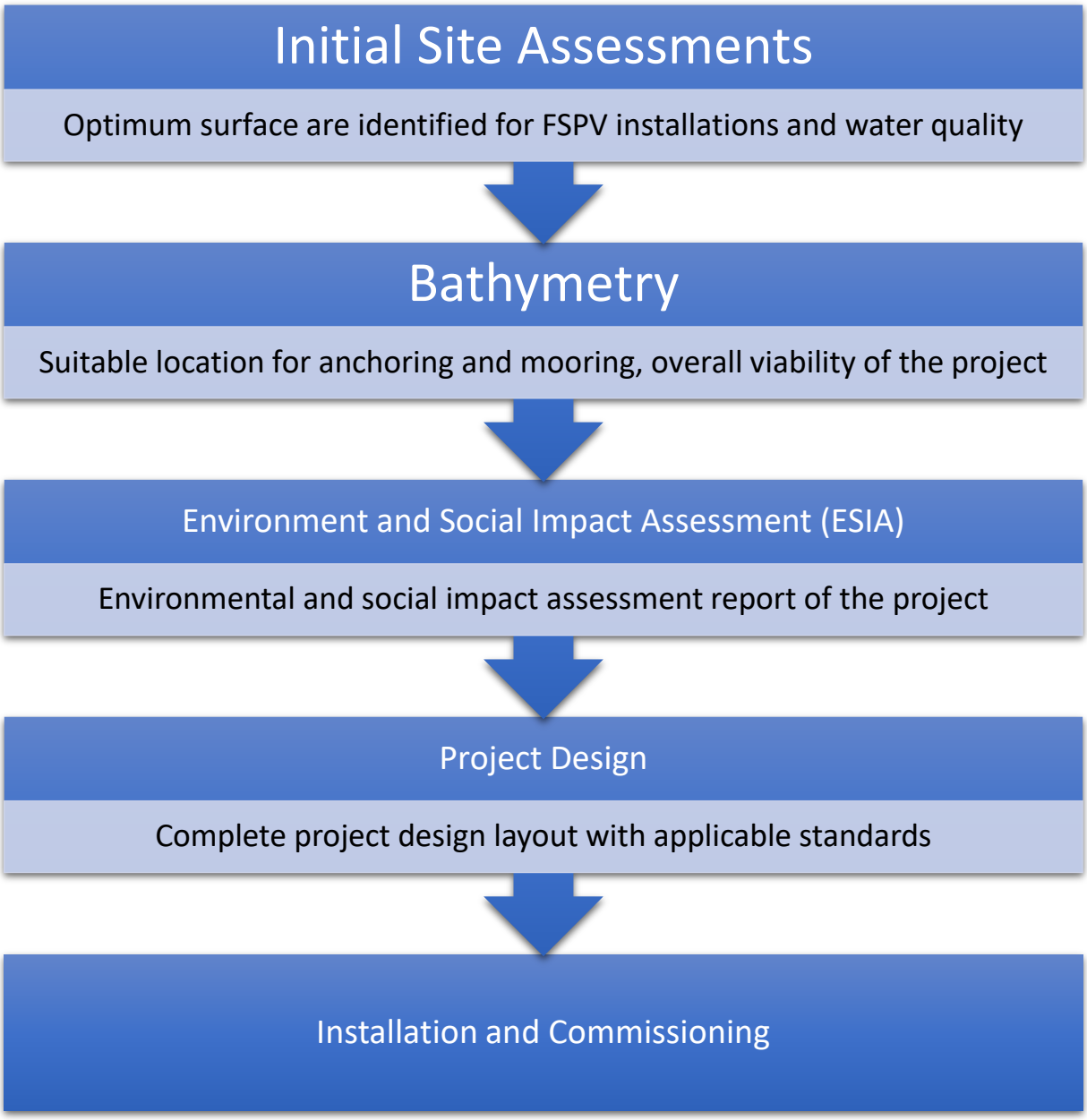
$$P_{inv} = P_a \times 1.25$$

Where P_{inv} is rated power of inverter, Pa is running power of the appliance.

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

IV. PROJECT GUIDELINES

Floating Solar Photovoltaic system as a technology is still in the evolving phase and involves several crucial elements including phases for establishing SPV plants that are not available publicly. Therefore, coming up with the most suitable practice guidelines and methods for the FSPC sector would undoubtedly be a great help to the project developers, investors and the policymakers in getting the idea and understanding the steps involved in the FSPV project development. In addition to this, FSPV plants are deployed in waterbodies and often subjected to harsh environments. Hence it is important to consider relevant standards to ensure the quality of the FSPV plant components while being attentive of not depreciate the quality of water as well as local bio-diversity. In a nutshell, major steps in developing an FSPV project can be understood through the flow chart of floating solar PV design as shown below:



V. CASE STUDIES

5.1 Banasurasagr Reservoir, Wayanadu, Kerala.

The floating solar power plant, was commissioned and inaugurated on 4th December 2017. The project will generate about 700,000 kWh per year which will be fed directly to the grid. Previous studies have proved that

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

floating solar panels generate higher output due to lower ambient temperature existing on the surface of the waterbody. Apart from this evaporation of water from the reservoir will be minimal as the surface is covered by solar panels.

- Location : Wayanad
- Latitude : 11.67° N
- Longitude : 75.96° E
- Altitude : 2423



Figure 3: Satellite view of Banasurasagr Reservoir

System Details

The 500 kWp Grid Interactive Floating Solar Power Plant in the Banasura Sagar dam, Wayanad is the first of its kind in India. The project is designed for Kerala State Electricity Board (KSEB) and the solar photovoltaic array is installed on 18 floating platforms made of Ferro cement floaters with hollow insides which are able to adapt to varying reservoir water levels by means of an innovative anchoring system. Floating solar farm floats on 6,000 square metres of water surface of the reservoir.

Upkeep and Maintenance

Dust and bird droppings were noticed in approximately all parts of the PV array. Frequent cleaning provisions are required to ensure that the array is clean and thereby ensuring maximum energy yield.

Earthing

Water Earthing: Supports are provided around the floating platforms to which the earthing plate is attached. The earthing plate is placed in that portion of the floating platforms which is submerged in water.

Armor Earthing: During occurrence of earth faults, the metallic sheaths are expected to carry a substantial proportion of the total fault current. For safety and reliable operation, the shields and metallic sheaths of power cables must be grounded. Armour earthing is done at the site which is good for the system safety.

VI. CONCLUSION

Solar energy is the most abundant source of clean energy over most parts of India. Utilizing FSPV systems, this energy can be harnessed on areas that have no competitive purpose and is often closer to areas of consumption compared to traditional power plants. In contrast to land mounted photovoltaic systems, the advantages of FPV power plants are less evaporation of water, greater efficiency and a reduction the emissions of CO2 greenhouse gas, which leads to the expanding of these systems in many countries. In countries with semi-arid and arid locations, water crisis is a huge problem and the use of FPV modules to reduce the evaporation rate of water is the right choice. In general, the sun in these countries is less expensive. Floating PV plant design includes all

aspects, even mechanical and electrical functions. In this case, this work describes the various possible configurations of the Floating PV grid connections and the use of multi-level DC-DC converters when connecting the FPV panels to the grid network. By moving towards FSPV integration into the city, FSPV systems encourage a faster and more economic development of solar projects, as it reduces the burden of land acquisition on the governments and encourages projects to be finished on time (due to the less probability of prolonged legal disputes regarding acquisition) by contractors. Given the spatial spread of the lakes, PV-EV-based charging stations can be established, reducing the payback period of the plants, enhancing accessibility for electric vehicle users in urban areas and encouraging start-ups, including addressing mobility issues using micro-mobility vehicles.

VII. REFERENCES

[1] World energy outlook 2020. 2020, International Energy Agency.

[2] Yoshikawa, K., et al., Silicon heterojunction solar cell with interdigitated back contacts for a photoconversion efficiency over 26%. 2017. 2(5): p. 1-8.

[3] Richter, A., M. Hermle, and S.W.J.I.j.o.p. Glunz, Reassessment of the limiting efficiency for crystalline silicon solar cells. 2013. 3(4): p. 1184-1191.

[4] Ndiaye, A., et al., Degradations of silicon photovoltaic modules: A literature review. 2013. 96: p. 140-151.

[5] Ziar, H., et al., Photovoltatronics: intelligent PV-based devices for energy and information applications. 2021. 14(1): p. 106-126.

[6] Lee, N., et al., Hybrid floating solar photovoltaics-hydropower systems: Benefits and global assessment of technical potential. 2020. 162: p. 1415-1427.

[7] Bellini, E., Japan's largest floating PV plant catches fire after Typhoon Faxai impact, in PV Magazine. 2019.

[8] Floating solar: Europe's largest power plant is French, https://www.planet.veolia.com/en/floating-solar-power-plant-piolenc-france,. 2019.

[9] Cazzaniga, R., et al., Floating photovoltaic plants: Performance analysis and design solutions. Renewable and Sustainable Energy Reviews, 2018. 81: p. 1730-1741.

[10] Floating in the Alps, PV magazine, November 2019. https://www.pv-magazine.com/2019/11/18/floating-in-the-alps/.

[11] https://oceansun.no/project/kyrholmen/.

[12] Energy Sector Management Assistance Program, Solar Energy Research Institute of Singapore, Where Sun Meets Water: Floating Solar Market Report. 2019, World Bank.

[13] Liu, H., et al., Field experience and performance analysis of floating PV technologies in the tropics. 2018. 26(12): p. 957-967.

[14] Ziar, H., et al., Innovative floating bifacial photovoltaic solutions for inland water areas. 2020.

[15] Rosa-Clot, M., Floating PV Plants. 2020: Academic Press.

[16] Trapani, K., M.J.P.i.P.R. Redón Santafé, and Applications, A review of floating photovoltaic installations: 2007–2013. 2015. 23(4): p. 524-532.

[17] Sahu, A., et al., Floating photovoltaic power plant: A review. 2016. 66: p. 815-824.

[18] Ranjbaran, P., et al., A review on floating photovoltaic (FPV) power generation units. 2019. 110: p. 332-347.

[19] Ziar, H., et al., A comprehensive albedo model for solar energy applications: Geometric spectral albedo. 2019. 255: p. 113867.

[20] Séférian, R., et al., An interactive ocean surface albedo scheme (OSAv1. 0): formulation and evaluation in ARPEGE-Climat (V6. 1) and LMDZ (V5A). 2018. 11(1): p. 321-338.

[21] Hasanuzzaman, M., et al., Global advancement of cooling technologies for PV systems: a review. 2016. 137: p. 25-45.

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

[22] Dörenkämper, M., et al., The cooling effect of floating PV in two different climate zones: A comparison of field test data from the Netherlands and Singapore. 2021. 214: p. 239-247.

[23] Abdelal, Q.J.I.J.o.L.-C.T., Floating PV; an assessment of water quality and evaporation reduction in semi-arid regions. 2021.

[24] Cazzaniga, R., et al., Integration of PV floating with hydroelectric power plants. 2019. 5(6): p. e01918.

[25] MOLINA, P.S., Cleaning robot for floating PV. PV magazine, 2021 available online [https://www. pv-magazine.com/2021/03/08/cleaning-robot-for-floating-pv/].

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

FRUITS/VEGETABLE CLASSIFICATION USING DEEP LEARNING AND IOT

Mr. Ashish Rajeshwar Kulkarni*1, Amit Gupta*2,
Ajit Kumar Yadav*3, Anand Kumar*4
*1,2,3,4Department Of Electrical Department Delhi Technological University Delhi, India.

ABSTRACT

The classification of fruits and vegetables according to their quality has a significant impact on their price, customers, and Industries. Analyzing the quality by humans is always subjective and varies from person to person. It also becomes quite difficult to classify and predict their quality if the quantity is huge. The previous research is unable to explain the complete process for sorting the fruits/vegetables according to their quality using both software and IoT. We are using the deep learning algorithm on our servers to predict the quality and simultaneously using the hardware connected with the IoT to classify them according to the quality in the real world. It tries to make use of strategies for extracting color, size, and other attributes from processing images by comparing them with the trained dataset's properties. In addition, a processed image is created after image processing, which marks the individual objects in the raw image by finding the contour and applying the watershed algorithm. We can classify the different fruits and vegetables using IoT Machine with an accuracy of up to 90%.

Keywords: Classification, Deep Learning, Iot, Application, Fruits/Vegetables Quality Analysis.

I. INTRODUCTION

Judging the quality of fruits/vegetables becomes quite difficult for buyers or sellers due to the huge quantity and subjectivity in prediction. We have to face this issue during the whole supply chain management of fruits/vegetables from farmers to the consumers. We need to classify it in various stages such as when farmers sell it to the whole sellers or government vegetable market, the whole seller sells it to industries or consumers, etc. This problem could be simplified using the application that uses neural networks and conveyor belt systems for classification. Its cost is cheaper compared to the conventional laser system technology used by the food processing industries. We are building the android applications which provide GUI so that could be easily used by the users. The application is used to send data to the server which uses the technologies such as node js for handling the request, storing data, and python scripts which are used for calculating individual objects and predicting their quality of it using the neutral network. The resultant information is returned to the android app which is further used to control the IoT using the android app. IoT is arranged in such a way that it can sort the fruits/vegetables with the help of a conveyor belt, servo motors, etc. Thes technique could be used in food processing industries to make the classification faster, cheaper, and more objective.

II. BACKGROUND

Fruit rotting has substantial economic consequences. It is estimated that roughly a third of the cost of the fruit is spent on decaying materials. Furthermore, the sale of fruits will be impacted because consumers believe that spoiled fruits are harmful to their health (Péneau, Linke, Escher, & Nuessli, 2009). The Internet of Things (IoT) is rapidly expanding and has the potential to become a massive data source. IoT has opened up new possibilities in a variety of fields, but it has also brought up new issues that must be addressed. The potential use of IoT in the food safety area has received little attention, hence the goal of this study was to address that gap. Recent efforts have been made to use computer vision and machine learning techniques to automate the classification of fruits based on their outer appearance or freshness. The majority of this research of P. Moallem, A. Serajoddin, H. Pourghassem Computer vision-based apple grading for golden delicious apples based on surface features Inf Process Agric, 4 (2017), K. Tan, W.S. Lee, H. Gan, S. Wang. Recognizing blueberry fruit of different maturity using histogram oriented gradients and color features in outdoor scenes Biosyst Eng, 176 (2018), however, graded using the usual strategy of feature extraction and machine learning approaches. the apples have been classified by using color, texture, and shape feature descriptors, namely Global Color Histogram, Color Coherence Vector (CCV), Local Binary Patterns (LBP), Complete Local Binary Patterns (CLBP),

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

and Zernike Moments (ZM). The extracted features were used individually as well as in combination to train and test the machine learning techniques with the highest accuracy of 95.9% using a combination of CCV, CLBP, and ZM. An artificial neural network (ANN) based framework was proposed in a recent study on detecting the ripeness quality of bananas utilizing multiple parameters such as color, brown spot development, and Tamura statistical texture [6]. The performance of the ANN model was compared to that of SVM, naive Bayes, KNN, decision tree, and discriminant analysis, among other techniques. The proposed method has the greatest overall recognition rate of 97.75 percent(%), according to the findings. Other fruits, such as blueberries, have used similar methods to determine ripeness and maturity levels. K. Tan, W.S. Lee, H. Gan, S. Wang Recognising blueberry fruit of different maturity using histogram oriented gradients and color features in outdoor scenes Biosyst Eng, 176 (2018), pp. 59-72.

Using deep learning techniques to construct a fruit classification and grading system is an alternative approach to dealing with the aforementioned issues. Deep learning-based models may automatically extract relevant characteristics and do not require operator involvement. This approach has been followed in some recent studies[1][2][3][4].

A pre-trained Inceptionv3 deep learning model was used to create a grading system for the apple data set, which can categorize the apples into four classes with a top 5 accuracy of 90%. This system graded 150 apples using a self-collected data set utilizing a transfer learning strategy using the Fruit 360 dataset.

The following are some of our contributions to this study:

• We suggested and designed a low-cost machine vision system capable of automating the visual assessment of fruits in real-time.

• In order to detect visual abnormalities in fruits for the agriculture domain, we created an end-to-end approach employing computer vision and Neural network models (individual and ensemble).

• The system may operate in both offline and real-time modes, accurately grading many instances of fruits in a single image or in real-time utilizing an efficient segmentation approach.

III. THE SYSTEMS DESIGN

It requires the implementation of both software and hardware for implementing it and finally linked them together to form an automatic system.

A. Software Implementation

Here we are building the frontend Application, backend Server, and Deep learning algorithm Python Script on the server. We are going to describe each implement and its interlinking in detail.

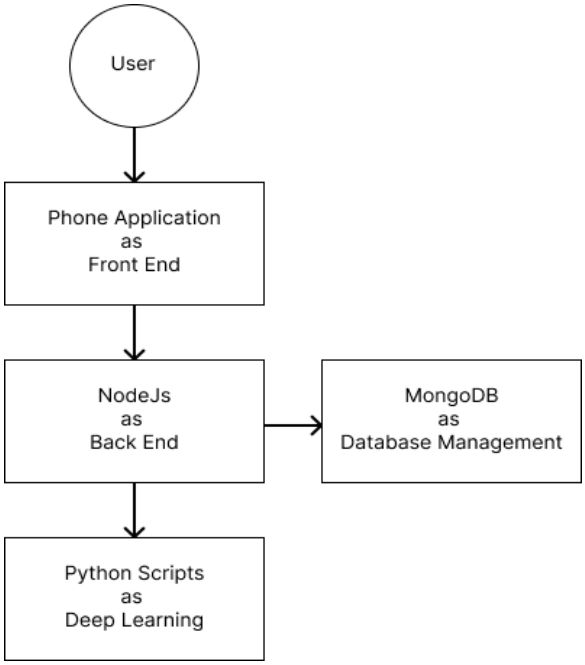


Fig 1: Block diagram of the interconnection of software applications

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

Frontend Application: We have built an android app whose main purpose is to interact with the User, send the user input data to the server, receive the results from the server and display it to the user.

At first, the user clicks the image of the fruits /vegetable using the camera and crops the required image out of it. Then, it is converted into a multipart part post request. Here we are using Retrofit Library for sending the Multipart request to the rest of the API on the server using the domain address. It is used to transfer/receive the file in original formats such as raw images and strings.

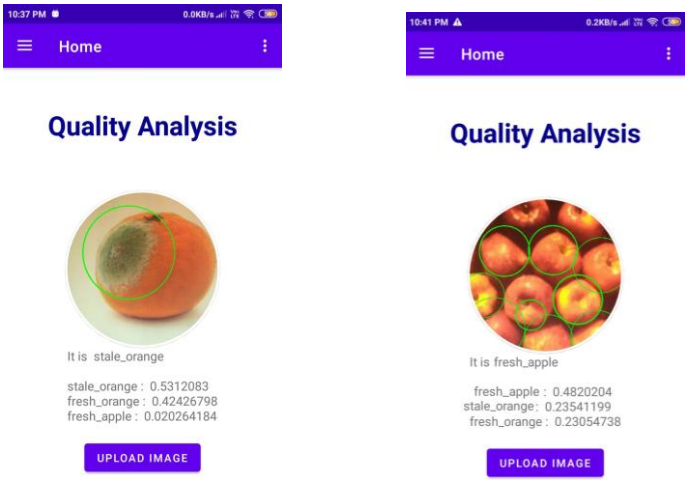


Fig 2: Frontend android application images

Then it waits till it receives the response from the server. When it receives the response then it shows it to the user and also sends it to IoT for the classification process.

BackEnd Server: Its purpose is to link the frontend app with the deep learning processes. It also handles the MongoDB DBMS for storing and retrieving the input and output data. It listens to its service of specific ports and routes. It saves the received image to a particular location. Finally, it calls deep learning scripts and passes the parameters.

Then it waits till the response comes from the deep learning python script. It converts the resultant processed image into the form of a base64 string so that it can be formatted in JSON form. In the end, it sends back the response result in the form of JSON data to the frontend app.

Deep Learning Process: It is a core process where we are using different algorithms to predict the quality analysis of fruits or vegetables.

There are certain algorithms that we are applying such as finding Contour, Neural Network, etc.

Contour Detection: At first we try to find out individual objects' outline boundaries So that we can apply the neural network to each object. The step that we use to find the contour of each object is given below:

Neutral Network: We use it for predicting the classification of fruits/vegetables in certain categories.

A neural network is a build-up of neurons, these neurons are the main processing part of the network. At 1st, we have the input layer which receives the input, and at last, we have the output layer which predicts the output of the Result. Between these two layers, we have several layers that are responsible for the computational part of the network.

Let's have an Image of 96X96 pixel = 9216 pixels as input. The neuron of this layer is connected to the neurons of another layer through the channel and each channel has weight and their submission acts as input to the next layer with the Bias Value. Its values are sent to the activation function which decides whose neuron value is transferred to the next layer and these neurons are known as the activation function. The process carries on to the output layer. Then we compare our result with the original result. Depending on the error magnitude backpropagation takes place and is weight-adjusted. Then the whole process is repeated until we get high accuracy.

We are using the Keras library for implementing training the dataset and building the model. Our dataset consists of more than 20 thousand images including custom and imported images of different categories.

We use this model for predicting the probability of the class of input image and then return its results to nodejs.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

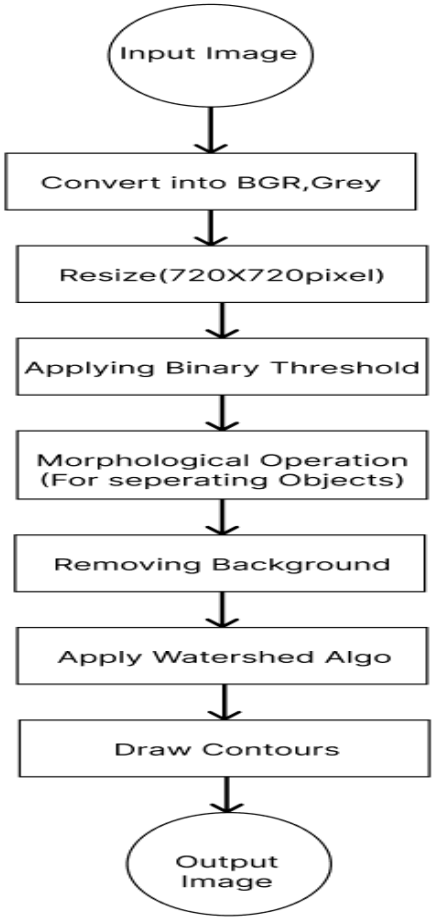


Fig 3: Block diagram of the implementation of contour algorithm

B. Hardware

In this part, we are integrating the Automatic Conveyor belt with the IoT devices to form a complete system. We have designed its 3D model to imagine what our project looks like.

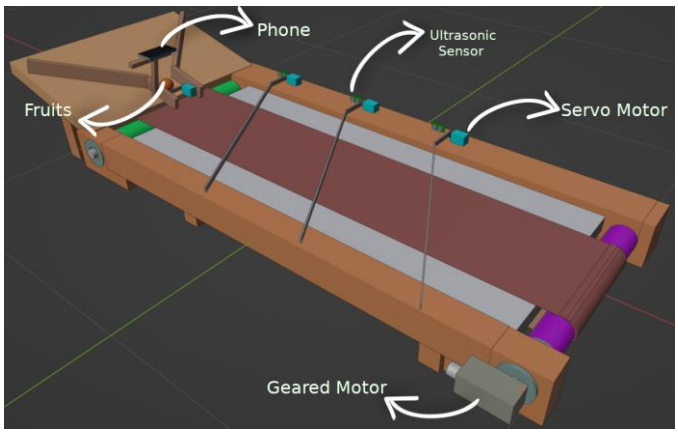


Fig 4: Blender image of the conveyor belt system

Mechanical Assemble: We are building a complete conveyor belt system for sorting or classifying fruits/vegetables in different categories. It involves the different steps of working.

1. **Initial stage:** At first, we put the different qualities of vegetables/fruits on the storage part of the system which is inclined at a 30-degree angle w.r.t. the ground. The phone camera captured the object which is at the starting position and sent the result to the ESP microcontroller which activated the servo 1 in the open state.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

2. Middle stage: The object at the last falls on the conveyor belt which consists of 2 rollers connected with the help of a belt. Out of which one is connected to a geared motor which helps in rotating the complete conveyor belt. So, the object which falls on the conveyor belt rotates at the same speed as the conveyor belt.

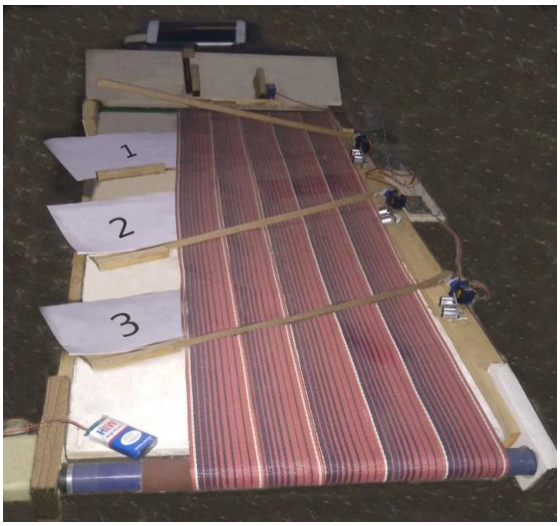


Fig 5: Hardware image of the conveyor belt system

3. Sorting stage: When it reaches the next servo position, it will be detected by the ultrasonic sensor one and the microprocessor will either close/open the path using the servo two depending upon the result. Suppose if it doesn't close then the object doesn't fall in category 1. It travels further on the conveyor belt to servo three positions where it is detected by the ultrasonic sensor 2. Suppose here the microprocessor closes the path using the servo 3 depending upon the result. Then it experiences two types of forces. First is the force due to the conveyor belt in the direction of motion and the other is due to the wooden Stick inclined at 30 degrees with respect to the conveyor belt. Due to the resultant of these forces, it will move toward point 2 as shown in fig.5 Therefore, it will fall into category two. Hence, One cycle of sorting is done. Then, the same process is repeated for other fruits/vegetables.

IoT Assemble:

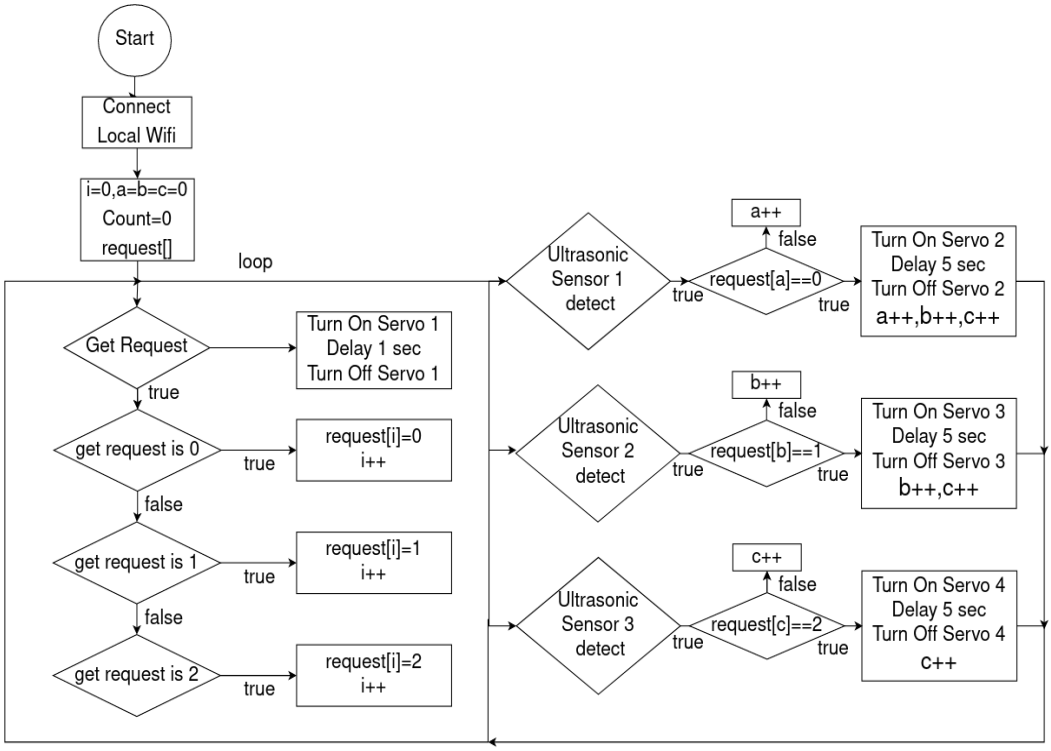


Fig 6: Flow chart of IoT

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

The node MCU handles the data received from the android application and controls the conveyor belt. For communicating the application, it connects to the local wifi and also sets all the servo motors in their default position.

Then, it starts a non-ending loop, in which we perform all the actions to classify the fruits/vegetables in their particular category. There are two parallel operations held during each loop. We discuss each in detail:

Loop1:In the first loop, we will store the results they receive from the application. For achieving this feature, whenever we receive any get request on node MCU, We store its received result value in a global dynamic Array. So that we can use it later to control the servo motor.

Loop2: We are using this loop to control the servo motor using the store result and the ultrasonic sensor detector. To achieve this, we are detecting whether it detects an object or not. For detecting objects using the HC-SR04 ultrasonic sensor which has a four-pin interface namely Vcc, ground, echo, and trigger, we will check if the distance of the object is less than 50 c.m. For finding the distance, we will send a high pulse to the trigger pin for 10 microseconds and calculate the duration it takes to reach the eco.

DistanceCm = duration * Sound_Velocity/2

where sound velocity is approximately 0.034 c.m./usec

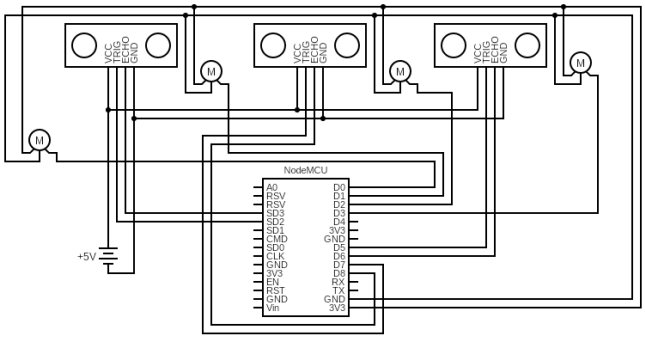


Fig 7: Circuit diagram of IoT

If the object is detected then it checks whether the object has a value in the resultant array equal to the corresponding position or not. Such as if the resultant array value for it is zero then we rotate the servo two to the zero degrees position with the help of pulse width modulation. The servo motor uses a feedback control system to adjust the shaft to an accurate position. Then the whole system waits for some time till the object has shifted its category. Then we rotate the servo to its original position. The same process takes place for other servo motors and ultrasonic sensors. In this way, we can send the objects to their correct position.

IV. RESULT

We are capable of performing the classification of fruit/vegetable with an accuracy of 90-95% with the help of a conveyor belt system. Also, with the help of our IoT conveyor belt, we are capable of sorting at the speed of 7-10 objects per minute.

Table 1: Measuring Classification Accuracy

S.No.	Sample	Name	Actual	Predicted
1.		Fresh apple	20	19
2.		Stale apple	20	18
3.		Unripe apple	20	18
4.		Fresh orange	20	18

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

S.No.	Sample	Name	Actual	Predicted
5.		Stale orange	20	19
6.		Fresh potato	20	18
7.		Stale potato	20	17
8.		Fresh banana	20	20
9.		Stale banana	20	19
10.		Fresh tomato	20	19
11.		Stale tomato	20	16

As per the above analysis, we got an accuracy of 91.36%. It could be further improved by providing more image data and training to the neutral network. The classification steps are shown in fig. 8.

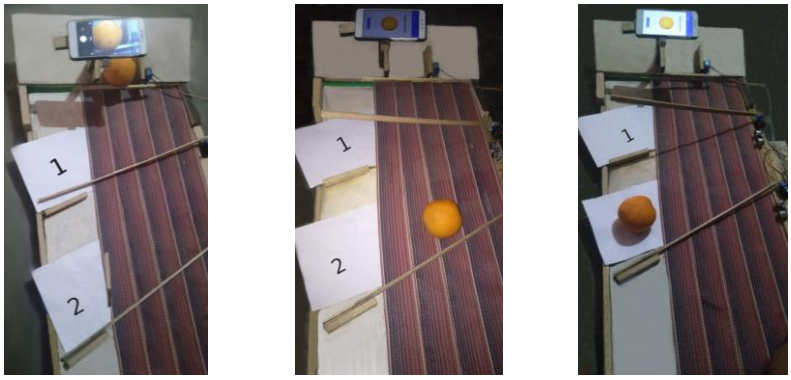


Fig 8: Hardware images showing the classification process

V. CONCLUSION

Using the deep learning algorithms with the IoT conveyor belt system provides a more convenient and cheaper solution for the classification of fruits and vegetables. An automatic system reduces the labor work in the sorting process of items. The user could directly communicate with the android application of the phone to initiate the process, which further communicates with the server for getting the response from the neutral network and control the IoT to perform the complete sorting process. The dataset of the neutral network consists of fresh, unripe, and rotten fruits/vegetables used for the implementation of it. The trained dataset provides an accuracy of 90% for judging the quality of fruits/vegetables. In our project, we can get the accuracy of approximately the same as the dataset accuracy. This study will have a great impact on judging the quality of fruits/vegetables in huge quantities by the industries or the buyers, automatically and objectively.

ACKNOWLEDGMENT

We would like to express our deep gratitude to the project supervisor Mr. Ashish Rajeswar Kulkarni for his passionate support, enthusiastic encouragement, and patient guidance throughout the project. We would also like to thank Delhi Technological University for providing us with a great learning environment and dedicated professors.

VI. REFERENCES

[1] S. R. Arlimatti, "Window Based Method For Automatic Classification Of Apple Fruit," Int. J. Eng. Res. Appl., Vol. 2, No. 4, Pp. 1010–1013, 2012.

[2] R. Szabó, A. Gontean, And L. Lie, "Cheap Live Color Recognition With Webcam," 2011 23rd Int. Symp. Information, Commun. Autom. Technol. Icat, Doi: 10.1109/Icat.2011.6102103.

[3] U. O. Dorj, M. Lee, And S. Seok Yun, "An Yield Estimation In Citrus Orchards Via Fruit Detection And Counting Using Image Processing," Comput. Electron. Agric., Vol. 140, Pp. 103–112, 2017,

[4] Doi:10.1016/J.Compag.2017.05.019

[5] M.S. Hossain, M. Al-Hammadi, G. Muhammad, "Automatic Fruit Classification Using Deep Learning For Industrial Applications" Ieee Trans Ind Informatics, 15 (2019), Pp. 1027-1034,

[6] Doi:10.1109/Tii.2018.2875149

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Green route synthesis and characterization of zinc oxide nanorods and its composite with PVA

Parisha Diwan, Kanika Sharma, Nitin K. Puri *

Nanomaterials Research Laboratory (NRL), Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110042, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
ZnO nanorods
Nanocomposite
Green synthesis
Osmunda Regilis
Polyphenolic compounds
Electrical properties

ABSTRACT

In this work, we have successfully synthesized Zinc Oxide Nanorods (ZnO-NR) via the Green Synthesis route. Green route has been incorporated to synthesize ZnO-NR and reduce the usage of toxic precursors as well as toxic by-products. Synthesis is carried out using Zinc Nitrate ($Zn(NO_3)_2{\cdot}6H_2O$) as a precursor and aqueous extract of Osmunda Regilis leaves. The plant extract of Osmunda Regilis leaves has a significant role as a reducing agent because of the presence of polyphenolic compounds. The synthesized ZnO-NR are then characterized by various characterization techniques- X-ray Diffraction (XRD), U-V Visible Spectroscopy (UV–vis), Fourier transform infrared spectroscopy (FTIR), Scanning electron microscopy (SEM). XRD results show that the as-synthesized ZnO sample is extremely crystalline, having a wurtzite structure. UV–vis absorption spectrum represents a typical spectrum for ZnO-NR which gives an absorption peak at 375 nm and exhibits a bandgap of 3.3 eV. Fourier Transform Infrared Spectroscopy (FTIR) shows the victorious chemical bond formation of ZnO nanorods. SEM micrographs display that formed Nanorods are Crystalline and have a lateral size of nanorods lying between 30 and 100 nm. Since it is challenging to form a film of bare powder for studying electrical properties therefore we formed ZnO/PVA composite thin film [8]. The electrical properties of ZnO/PVA nanocomposite thin films are studied which exhibit the ohmic behavior.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

1. Introduction

Nowadays nanotechnology is a basic need in every research field [1]. The interest of researchers in nanotechnology has been growing day by day. It is one of the important subjects which play a pivotal role in branches of applied physics and applied chemistry. It deals with the matter at different dimensions ranging from 1 to 100 nm. Nanomaterials have vast applications due to their size, shape, and structure. Nowadays the great focus has been shifted towards nanomaterials which show superior electrical, structural & optical properties [2].

Today nanomaterials are in demand because of their reduced size, exceptional surface active sites, and they are progressively being used in various sectors such as biomedical, drug delivery, and photothermal therapy due to their antimicrobial and anticancer properties [3]. Various types of nanomaterials have been extensively used such as metal oxides (ZnO, PdO, PTO) [4], metal nitrides (h-BN, AlN, GaN) [5], Transition Metal Dichalcogenides (TMDs –MoS_2, WS_2, $MoSe_2$, WSe_2). Among them, the great focus is on metal oxides because of their major properties such as optical, electrical, mechanical, and piezoelectric and it is considered the major factor for productive interaction with target molecules.

Among several metal oxides, ZnO is considered one of the versatile semiconductors which act as a transparent conductive oxide that opens a lot of applications [4]. The interesting properties of ZnO when coated on conducting substrate exhibits appreciable efficiency for solar cell and is depicted by the fabrication of quantum-dot sensitized solar cells [6]. Also, it has a large bandgap of about (3.2–3.8 eV) and crystallite size of (8–30 nm). It displays optical and luminescent properties in UV–Visible spectroscopy.

For the synthesis of ZnO nanoparticles, various methods are used like the Sol-gel route, Hydrothermal, Precipitation method, Chemical Vapour Deposition, Green route synthesis method, etc [1,6]. These types of approaches require high energy and high temperature and it involves poisonous chemicals which affect biological methods badly. Among them, the green approach is considered

* Corresponding author.
E-mail address: nitinkumarpuri@dtu.ac.in (N.K. Puri).

https://doi.org/10.1016/j.matpr.2022.04.491
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

Please cite this article as: P. Diwan, K. Sharma and N.K. Puri, Green route synthesis and characterization of zinc oxide nanorods and its composite with PVA, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.491

one of the advantageous approaches due to the eco-friendly approach also they are single step, safe, and economical and it has a lot of applications in farming, medical, and industries [7]. Due to the physicochemical nature of plant-based nanoparticles, this method provides an additional advantage of a high lifetime of nanoparticles that prevail over the disadvantages of conventional methods i.e. physical and chemical methods of nanoparticle synthesis. Also, the biological approach includes natural things which act as capping agents and stabilizing agents [8].

Here, we describe the green synthesis of ZnO nanorods using plant extract of Osmunda Regilis and Zinc salt (Zinc Nitrate) as a precursor. Osmunda regilis commonly known as Royal Fern is a species of deciduous fern with medical properties local to Africa, Europe, and Asia which is growing on the seaboard. It belongs to the family Osmundaceae. Its configurations are rounded in shape, moisture-loving, and natural. It accommodates various phytochemicals which show the presence of flavonoids, phenols, and triterpenoids and acts as a reducing agent. They have metabolites and some properties which have the capacity that can convert metal ions into nanoparticles. Several substances like sugars, alkaloids, phenolic acids, and polyphenols can reduce metal ions reduction value or strength. It is a single-step procedure and that is why it needs less amount of energy. It can control the growth of the crystal and the size of the particle because they act as a capping agent. It is done under the presence of ambient temperature, pressure, pH value which reduces the toxicity of chemicals [9]. To study the electrical properties of ZnO, a nanocomposite of ZnO/ PVA has been made [10,11].

2. Experimental section

2.1. Chemical reagents

The chemicals used in this synthesis are Zinc nitrate hexahydrate ($Zn(NO_3)_26H_2O$), Sodium hydroxide pellets (NaOH), ethanol (C_2H_5OH), and acetone (C_3H_6O), and Osmunda Regilis leaves extract which has a chemical composition of Hexahydrofarnesyl acetone ($C_{18}H_{36}O$), 2-4-di-t- butyl-phenol ($C_{14}H_{22}O$), and phytol ($C_{20}H_{40}O$). polyvinyl alcohol (C_2H_4O). All the chemicals are purchased from Fisher Scientific and Sigma Aldrich. Ultrapure Milli-Q water is used for washing and making solutions needed in the experiment.

2.2. Instrumentation

An X-ray diffractometer of the Bruker Advance D-8 model has been used to investigate phase identification. It produces Cu-K radiation with a wavelength of 1.5418 Å having a range from 10–80°. The use of a PerkinElmer Spectrum with a scanning range of 4000–400 cm^{-1} validates the presence of functional groups and the names of chemical bonds in the as-synthesized product by the peaks detected. The UV–Vis spectroscopy is conducted using Perkin Elmer750. SEM micrographs are obtained from EVO40.

2.3. Preparation of leaves extract

Garden-fresh Osmunda Regilis (Royal Fern) leaves were extracted from the Delhi Technological University campus, which was firstly washed with tap water and then brushed up using deionized water twice or thrice; after that leaves were dried and then ground. A few amounts of Osmunda Regilis powder is dissolved in a few amount of DI water. The solution was then boiled at 60 °C for 20 min and then it got filtered through Whatman filter paper after filtration it turns brown as shown in the Fig. 1 and then stored in a cold and dry place for further use in the future (see Figs. 2–4).

2.4. Synthesis of ZnO nanorods

In the green route of ZnO NPs by zinc nitrate, 3.51 mL of 0.05 M [$Zn(NO_3)2.6H_2O$] solution was mixed with 38 mL of Osmunda Regilis (Royal Fern) aqueous leaves extract. The mixture was then stirred for 2 h at 400 pm. Then 0.02 M NaOH was added dropwise to reach pH 12 and ultrasonication it stepwise so that all the particles get dispersed. The mixture was again stirred for an extra 1 h at 250 rpm till a solid product with a light yellow color was formed. The precipitate was purified by some re-dispersions from deionized water or ethanol and after that, the product was centrifuged at 6000 rpm for 10 min in two go. The end result was a white powder material that had been dried overnight.

2.5. Preparation of ZnO/PVA thin film

The ZnO /PVA Nanocomposite films were made by the drop-casting method. Synthesized ZnO was added in weight percent ratio by wt% of 5 to PVA from the formula as given below:

$$a\,(wt\%) = 100 \times \frac{wz}{wz + wp}$$

where w_z and w_p represent the weight of ZnO and PVA.

The Nanocomposite is produced by dispersing ZnO powder into PVA (2 g) which is dissolved in 50 mL of DI water and then the solution is kept for stirring for 1 h at 80 °C to prevent it from any type of agglomeration. Then films were made through the drop-casting method. Lastly, the films were dried in an oven at 80 °C.

3. Results and discussion

3.1. X-ray diffraction

XRD pattern of synthesized ZnO nanorods which represents the structure of the nanoparticles that we synthesized is crystalline. The keen diffraction peaks remarked at 2θ values which are confirmed from **(JCPDS card no-00-001-1136)** 31.88°, 34.52°, 36.36°, 47.56°, 56.64°, 62.96°, and 66.6°. The peaks are listed as (1 0 0), (0 0 2), (1 0 1), (1 0 2), (1 1 0), (1 0 3), (1 1 2), and (2 0 2) diffraction lattice planes which verify the hexagonal wurtzite structure for the ZnO nanorods we synthesized. The crystalline size i.e. t is found using Debye Scherer's formula-.

$$d = \frac{0.9\lambda}{\beta Cos\theta}$$

where λ is the wavelength of X-ray (1.54 Å), β is the value of Full-width half maxima (radian), θ is Bragg's angle [2dSinθ = n λ].

Sample	d-spacing (nm)	FWHM (radian)	Estimated Crystalline Size (nm)
1.	11.52	0.227	37.4

3.2. U-V Visible spectroscopy

UV–visible spectra of synthesized zinc oxide nanorods are shown in the above figure. The peak is at around 375 nm because of high binding energy at normal/room temperature. Also, we know that by absorption spectroscopy the bandgap increases when

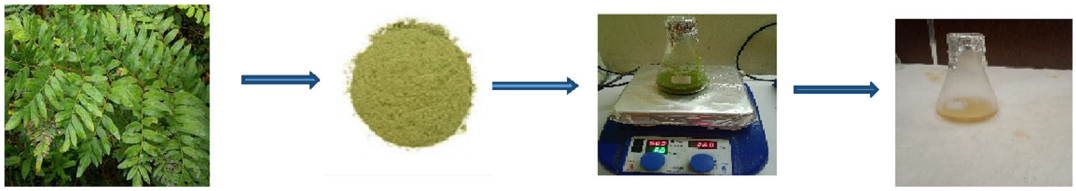


Fig. 1. Schematic diagram representing preparation of plant extract.

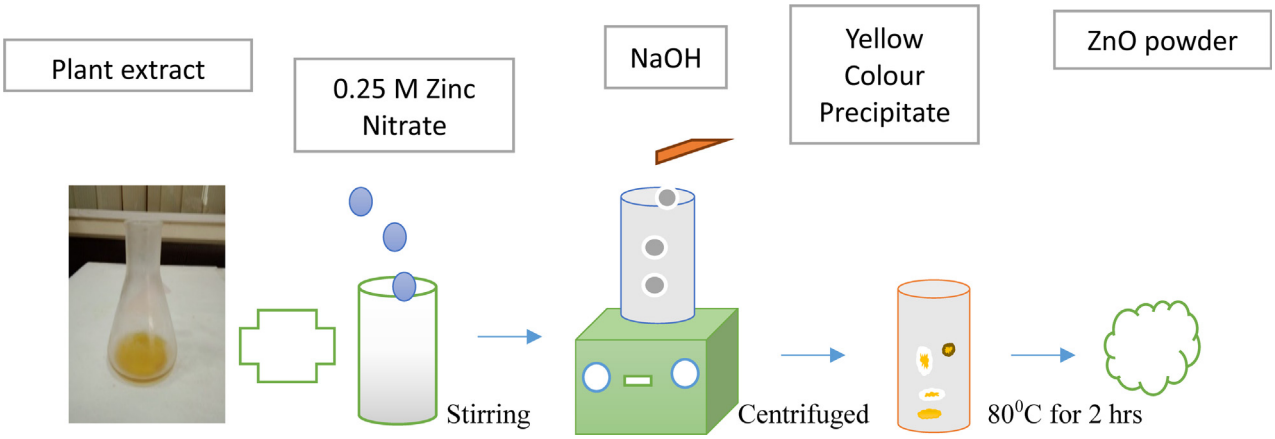


Fig. 2. Schematic representation of the synthesis of ZnO Nanorods.

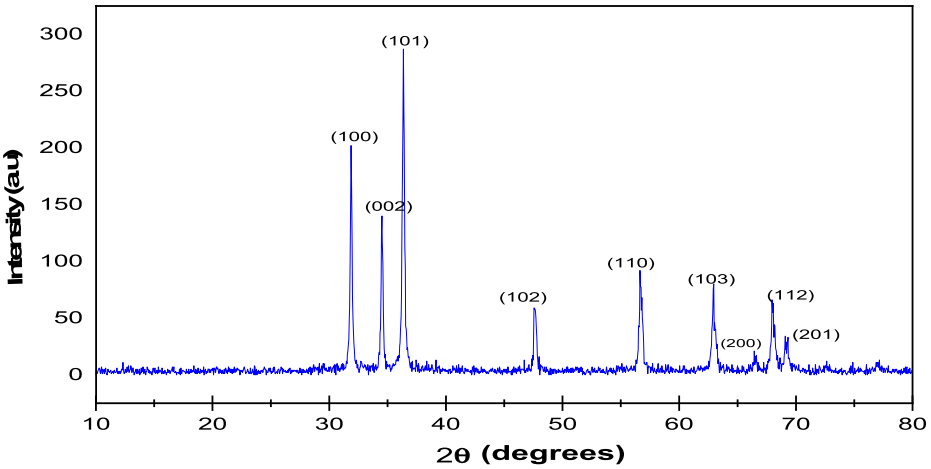


Fig. 3. Graphical depiction of XRD Spectra of prepared ZnO Nanorods.

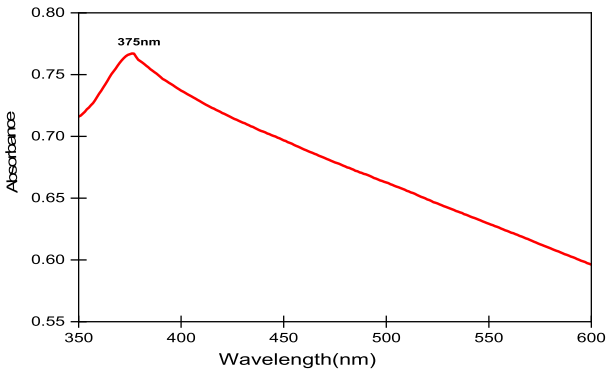


Fig. 4. Graph representing U-V Visible spectra of ZnO nanorods.

the size of the particle decreases as ($t \propto 1/E_g$) where E_g is the band-gap gap of ZnO NPs calculated from the above formula-.

$$(\alpha E)^{1/n} = \beta(E - E_g)$$

where the power factor of the transition mode is n, E_g is the optical bandgap energy, the band tailing parameter is β, and E is calculated from the above formula i.e. E = hν, where h is the Planck's constant, ν is the frequency of UV – Vis spectra. ZnO is a direct bandgap material for allowed transitions therefore n = 1/2.

From the above formula, the bandgap is calculated to be 3.2 eV. Whereas, the actual band gap for ZnO nanorods is found to be 3.8 eV which is larger than the bandgap we found this is how we see that the particle size is inversely proportional to the bandgap because if the particle size is less than it enlarges the energy bandgap.

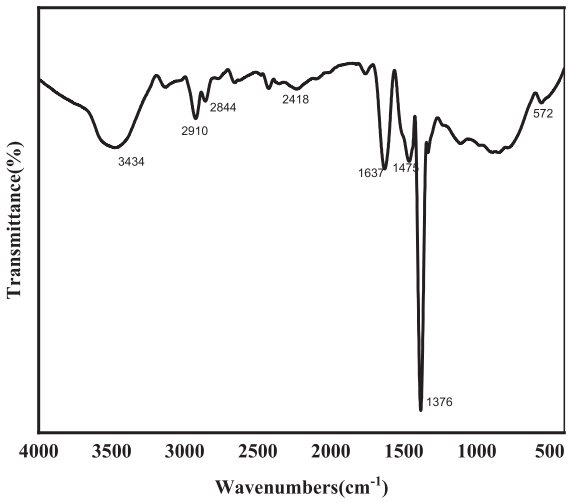


Fig. 5. FTIR spectra of ZnO nanorods.

3.3. Fourier transform infrared spectroscopy

FT-IR studies of Osmunda Regilis aqueous leaves extract and product of synthesized ZnO-NR as shown in Fig. 5. Infrared studies were achieved to find out the purity, nanoparticles nature, and also the proximity of phytochemicals in the material. The phytochemicals such as phenols, amines, alcohols and carboxylic acids can interact with the zinc surface and support in the maintenance of ZnO-NR. The peak at 1196cm^{-1} specifies that it is a C—N stretch which specifies that there is a presence of amines. The peak at 1637 cm^{-1} corresponds to Zn-O stretching and deformation vibration. The peak at 3434 cm^{-1} indicates the OH stretching vibrations which indicate there is the presence of the carboxylic acid functional group.

3.4. Scanning electron microscopy

Fig. 6 shows the SEM images of ZnO Nps, which clearly shows the synthesized ZnO Nanoparticles are nanorods and their size lies between 30 and 100 nm having a crystalline structure, high magnification Fig. 6a and low magnification (6b) SEM images are taken. The image with low magnification shows uniform ZnO nanorods with uniform yield. The image with high magnification nanorods surface is plane. Very minute second particle growth components are interposed (see Figs. 7–8).

3.5. Electrical properties of ZnO/PVA nanocomposite thin films

From the I-V curve, we obtained of ZnO/PVA nanocomposite thin film were measured at normal temperature as given in Figure, we can easily demonstrate that the graph obeys ohms law from its linearity above 0.5 V.

As ZnO represents semiconducting behave as PVA i.e Polyvinyl alcohol represents conducting nature which represents its better charge transport through it [3]. This linear nature in the interface of metal and insulators is very salient in the growth of electronic devices with better performance and trustability.

4. Conclusions

In this exertion, ZnO-NR was synthesized by Osmunda Regilis leaves aqueous solution and Zinc Nitrate, with NaOH as a precursor. The results of XRD committed the regulation of the synthesis procedure which indicates the creation of crystallite zinc Nanorods along with a hexagonal wurtzite structure. UV–Visible spectrum showed a clear peak at 375 nm, from which we determined the value of band gap i.e 3.3 eV. The average size of synthesized ZnO-NR was 37.4 nm. FT-IR studies show the formation of ZnO-NR and contain various phytochemicals which we can determine from peaks at different positions which act as capping and reducing agents for the synthesized ZnO-NR and the I-V characteristics of ZnO/PVA nanocomposite thin film reveal that it is Ohmic nature.

CRediT authorship contribution statement

Parisha Diwan: Experimental work, Methodology, Writing - original draft, Software, Conceptualization, Methodology. **Kanika Sharma:** Conceptualization, Formal analysis, Writing – review &

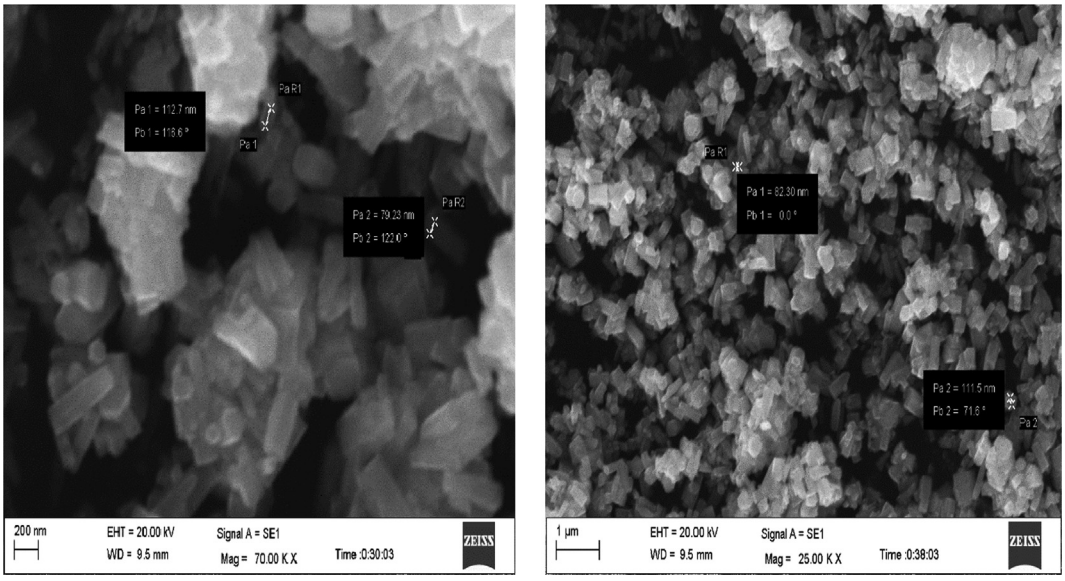


Fig. 6. SEM images of ZnO Nanorods (ZnO-NR) (6a) SEM images showing the particles at a scale of 200 nm and (6b) SEM images showing the particles at a scale of 1 µm 200 nm.

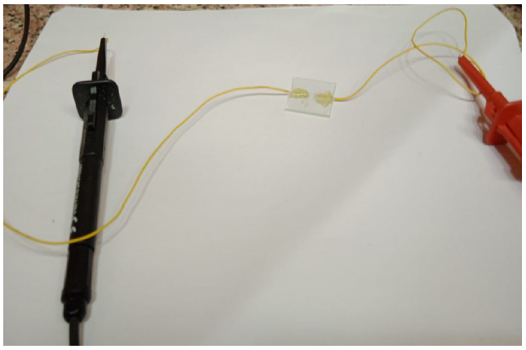


Fig. 7. Electrical measurements of ZnO/PVA thin film.

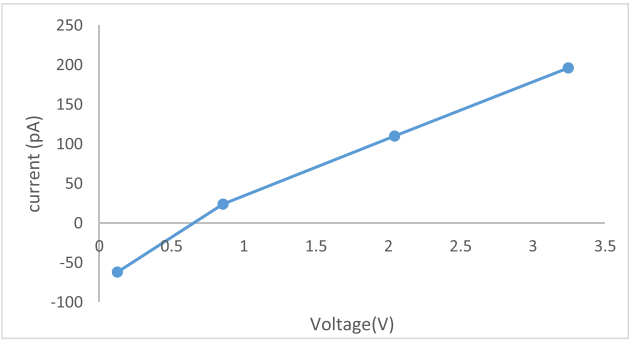


Fig. 8. I-V curve of ZnO-NR/PVA Nanocomposite thin films.

editing, Data curation. **Nitin K. Puri:** Supervision, Project Administration, Resources, Writing – review & editing, Validation.

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.

Acknowledgement

The authors are greatly thankful to Prof. Jai Prakash Saini, Vice-chancellor, Delhi Technological University, Delhi, India for arranging the research facilities. We are thankful for the cooperation rendered by Sandeep Sarpal for their support in the analysis of characterization techniques.

References

[1] S. Fakhari, M. Jamzad, H. Kabiri Fard, Green synthesis of zinc oxide nanoparticles: a comparison, Green Chem. Lett. Rev. 12 (1) (2019) 19–24.
[2] C.L. David, B.C. Bruce, C. Gene, P. Seong-Ju, R.M. Gary, "Electrical properties of ZnO." Zinc Oxide—A Material for Micro-and Optoelectronic Applications, Springer, Dordrecht, 2005, pp. 37–46.
[3] N. Minha, A. Usman, K. Bushra, C. Bin, Green route to synthesize Zinc Oxide Nanoparticles using leaf extracts of Cassia fistula and Melia azedarach and their antibacterial potential, Scientific Rep. 10 (1) (2020) 1–10.
[4] H. Hashim, S.S. Shariffudin, P.S. M. Saad, H.A.M. Ridah, Electrical properties of ZnO thin films prepared by sol-gel technique, in: 2010 International Conference on Electronic Devices, Systems and Applications. IEEE, 2010.
[5] K. Sharma, N.K. Puri, Enhanced electrochemical performance of hydrothermally exfoliated hexagonal boron nitride nanosheets for applications in electrochemistry, J. Electrochem. Soc. 168 (5) (2021) 056512.
[6] Md Jahidul Haque, Md Masum Bellah, Md Rakibu Hassan, Suhanur Rahman[1], Synthesis of ZnO nanoparticles by two different methods & comparison of their structural, antibacterial, photocatalytic and optical properties, Nano Express 1.1 (2020) 010007.
[7] A. Diallo, B.D. Ngom, E. Park, M. Maaza, Green synthesis of ZnO nanoparticles by Aspalathus linearis: structural & optical properties, J. Alloys Compd. 646 (2015) 425–430.
[8] A.N. M. Nurul, A.A. Nur, J. Junaidah, H. Abdul, Plant extract as a reducing agent in the synthesis of metallic nanoparticles: a review, Adv. Mater. Res. Vol. 832. Trans Tech Publications Ltd, 2014.
[9] M. Maruthupandy, M. Anand, G. Maduraiveeran, S. Suresh, A.S.H. Beevi, R.J. Priya, Investigation on the electrical conductivity of ZnO nanoparticles-decorated bacterial nanowires, Adv. Nat. Sci.: Nanosci. Nanotechnol. 7 (4) (2016) 045011.
[10] Q.K. Muhammad, K. Davood, N. Nazish, S. Amir, H. Tanveer, K. Zeeshan, Z. Chunhong, K.I. Soo, Preparation and characterizations of multifunctional PVA/ZnO nanofibers composite membranes for surgical gown application, J. Mater. Res. Technol. 8 (1) (2019) 1328–1334.
[11] M.K.D.D. Andualem, A.Z. Enyew, M.H.C. Ananda, Synthesis of ZnO and ZnO/PVA nanocomposite using aqueous MorinOliveiraira leaf extract template: antibacterial and electrochemical activities, Rev. Adv. Mater. Sci. 59 (1) (2020) 464–476.

Applied Biochemistry and Biotechnology
https://doi.org/10.1007/s12010-022-03931-7

ORIGINAL ARTICLE

Highly Sensitive Enzymatic Biosensor Based on Polyaniline-Wrapped Titanium Dioxide Nanohybrid for Fish Freshness Detection

Deeksha Thakur[1] · **Chandra Mouli Pandey**[1,2] · **D. Kumar**[1]

Accepted: 20 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Herein, we envisage the fabrication of a highly sensitive enzymatic electrochemical biosensor for selective detection of xanthine (Xn) using xanthine oxidase (XOs) immobilized polyaniline-wrapped titanium dioxide (PANI@TiO_2) nanohybrid as a sensing platform. The PANI@TiO_2 nanohybrid was synthesized via chemical polymerization using ammonium per sulfate as an oxidant. Various microscopic, spectroscopic, and electrochemical techniques have been utilized to confirm the electrophoretic deposition of the PANI and PANI@TiO_2 nanohybrids on to indium tin oxide (ITO) coated glass substrate. The fabricated XOs/PANI@TiO_2/ITO electrode exhibits enhanced electron transfer kinetics with an electron transfer rate constant of 0.904 cm s^{-1}. The electrochemical results show that the fabricated biosensor can detect Xn in the concentration range 1–100 μM, with a limit of detection of 0.1 μM ($S/N=3$) and a response time of 10 s. The validation of the biosensors has been conducted using real samples obtained from the *rohu* (*Labeo rohita*) fish. The proposed biosensor can be a reliable analytical tool for determining Xn concentration in commercial fish and meat samples.

Keywords Xanthine · Electrophoretic · Conducting polymers · Titanium dioxide · Biosensor

Introduction

There has been a significant increase in demand for fish as a nutritious and healthful food, and this has created an urgent need to detect its freshness [1]. After fish death, various autocatalytic reactions, decomposition of nucleotides, start immediately by

✉ Chandra Mouli Pandey
cmp.npl@gmail.com

✉ D. Kumar
dkumar@dce.ac.in

Deeksha Thakur
thakur.diksha2@gmail.com

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

[2] Department of Chemistry, Faculty of Science, Shree Guru Gobind Singh Tricentenary University, Gurugram 122505, Haryana, India

Published online: 06 May 2022

endogenous enzymes that are present in the intestine and muscle tissues [2]. The rapid degradation of adenosine triphosphate (ATP) results in increased concentration of inosine, which by bacterial metabolism is transferred to hypoxanthine and xanthine (Xn) [3]. Xn is a by-product of the degradation of nucleotides and its concentration increases with storage that causes a bitter taste in fish meat [4]. Xn is found as a constituent of nucleic acids and it accumulates not only in fish but also in human organs (heart, kidney, and skeletal muscle). The excess volume of Xn concentrations in human blood and urine may lead to various physiological issues such as gout, hyperuricemia, and preeclampsia [5, 6]. Therefore, there is an urgency for the detection of Xn as it is an indicator for fish freshness. The traditional methods commonly used for Xn determination are based on spectrophotometric and chromatographic techniques that require sample pretreatment, are time consuming, laborious, and may result in toxic waste [7]. The development of a quick, accurate, reliable, and cost-effective technique for determining fish freshness is a major requirement in the marine food industry [8–10]. The new analytical strategies may lead to the development of assay methods with anticipated reusability, portability, reversibility sensitivity, selectivity, simplicity, and the possibility to monitor a variety of parameters simultaneously. In this context, electrochemical biosensors can be used as promising alternative methods for faster, reliable, and simpler ways for detection of Xn [11, 12].

Conducting polymers (CPs) have aroused much research interest in the fabrication of biosensors because of their excellent physicochemical and electrical properties that can be tailored to specific needs. Electrochemical biosensors based on CPs are more affordable, simple to fabricate, and offer a direct electrical readout for recognizing natural analytes with exceptional sensitivity and selectivity [13]. Among the different CPs, polyaniline (PANI) has been widely used for the development of biosensors because of its ease of processing, long-term environmental stability, and functionality-rich chemical structure [14, 15]. Inspite of these advances, the main drawback of using PANI-based biosensors is their low response, low detection limit, and inadequate selectivity. Further, to improve the desired functional and structural properties of PANI such as high aspect ratio, enhanced mechanical strength, and electrical properties, it can be integrated with nanostructured metal oxides (NMOs). These NMOs provide a large surface area and multiple active sites for the immobilization of the biomolecules, which result in the efficient transfer of the electrons between the electrode and the electrolyte [16]. In this context, titanium dioxide (TiO_2) nanoparticles have aroused a lot of attention among various NMOs because of their exceptional properties such as outstanding surface-to-volume ratio, chemical stability, biocompatibility, ability to immobilize biomolecules, and exceptional thermoelectric capabilities [17–21]. The integration of PANI with TiO_2 nanoparticles can minimize the agglomeration of the nanoparticles, thus, improving the stability and allowing uniform dispersion in the polymer matrix [22, 23].

In this work, PANI@TiO_2 nanohybrid has been chemically synthesized using the oxidative polymerization technique. The synthesized PANI@TiO_2 nanohybrid has been used as an electrode material for the detection of Xn to evaluate fish freshness by depositing it electrophoretically onto indium tin oxide (ITO)-coated glass substrate. The validation of the fabricated biosensor was investigated in the real sample by determining the Xn concentration in the fish meat. The unique architecture of PANI@TiO_2 nanohybrid provides a large electro-active surface area, fast electron kinetics, and improved catalytic activity leading to enhanced electrochemical performance [16].

Materials and Methods

Chemicals

Titanium tetrachloride ($TiCl_4$), aniline (99.5% pure), hydrochloric acid (HCl, 0.5 M), and ammonium persulfate (APS, $\geq$ 98% pure) have been procured from Fisher-Scientific, India. Glutaraldehyde, xanthine oxidase (XOs, $\geq$ 7 units/mg protein), xanthine ($\geq$ 99.5%), and other reagents and solvents were purchased from Sigma-Aldrich, USA. The solutions were prepared in double-distilled water (Milli-Q, Millipore, 18.2 MU) and the glasswares were autoclaved before use.

Synthesis of TiO_2 Nanoparticles

For nanostructured TiO_2, 1 mL of $TiCl_4$ was added to dilute H_2SO_4 (10%) solution in an ice bath with constant stirring. The solution turns grey due to the hydrolysis of $TiCl_4$. It was then held at 60 °C for heating until a clear solution was observed, and then aged for 1 h at room temperature. After that, 40 mL (drop by drop) of NH_4OH was added to the solution until the pH reached 7.0 and the solution turned white. The solution was allowed to cool and after that, it was washed and rinsed using DI water. The formed precipitate was vacuum dried and calcined at 400 °C, yielding an off-white tinted powder of TiO_2 nanoparticles [24, 25].

One-Pot Synthesis of PANI@TiO_2 Nanohybrid

The PANI@TiO_2 nanohybrid has been synthesized via a chemical approach using aniline as the monomer and APS as the oxidant. One milliliter of distilled aniline was dissolved in 0.5 M HCl in a 100 mL beaker. During the process, the beaker was set aside in an ice bath at a temperature of − 2 °C to − 4 °C for 10 h under constant stirring. After that, 1 mL of APS was added to the solution along with 0.2 mg of synthesized TiO_2 and kept in an ultrasonication bath for further dissolution. After 2 h of sonication, the solution was left undisturbed overnight at 4 °C. Further, the solution was rinsed and filtered using 2 M of HCl, and the resultant green-colored PANI@TiO_2 nanohybrid was dried at 60 °C for 24 h [26]. The same method was used to synthesize PANI without TiO_2 doping.

Electrophoretic Deposition (EPD) of PANI@TiO_2 Nanocomposite

The synthesized PANI@TiO_2 nanohybrid was deposited using the electrophoretic deposition (EPD) technique (Genetix, Model GX300C) onto a pre-hydrolyzed ITO electrode using a two-electrode system where platinum was taken as a counter-electrode. Both the electrodes were placed at a distance of 0.5 cm. The PANI@TiO_2 nanohybrid was sonicated in deionized water until a clear solution was obtained. For the deposition, 10 mL of the sonicated solution was diluted with 5 mL of ethanol and subjected to EPD. The deposition of the film at various voltages was carried out and the smooth film formation

was observed at 10 V and the time was 25 s. After the deposition, the prepared electrode was kept for drying.

Fabrication of PANI@TiO_2 Nanohybrid-Based Biosensor

For the immobilization of XOs, the PANI@TiO_2/ITO electrode was treated with 0.1% glutaraldehyde and left for 1 h at ambient temperature for the activation of the functional group present on the electrode surface [1]. Further, the electrode was washed with phosphate buffer saline (PBS; 100 mM, pH 7.4, 0.9% NaCl), followed by the covalent immobilization of XOs (20 μL). The electrode was kept overnight at room temperature in moist condition. The scheme for the synthesis of the PANI@TiO_2 nanohybrid and fabrication of the electrode is shown in Fig. 1.

Instrumentation

Bruker D-8 Advance X-ray diffractometer with monochromatic (Cu Kα) radiation ($\lambda = 1.5406$ A°) was used to record the X-ray diffraction (XRD) pattern in the 2θ range (10–80°) at a scan speed of 2°/min. To investigate the formation of PANI@TiO_2 nanohybrid, FT-IR measurements were performed using the Nicolet™ iS10 at a frequency range of 400–4000 cm^{-1}. Thermal studies were carried out to investigate the thermal degradation of the synthesized material in the temperature range of 25 °C to 1000 °C at the rate of 10 °C/min in a nitrogen environment using thermogravimetric analysis (TGA; Perkin-Elmer 4000). Scanning electron microscope (SEM, FEI NOVA NANOSEM 650) was used to analyze the surface morphology of PANI@TiO_2/ITO and XOs/PANI@TiO_2/ITO electrodes. The Autolab potentiostat/galvanostat (Eco-Chemie, the Netherlands) having a three-electrode system was used to perform electrochemical experiments with platinum as an auxiliary electrode

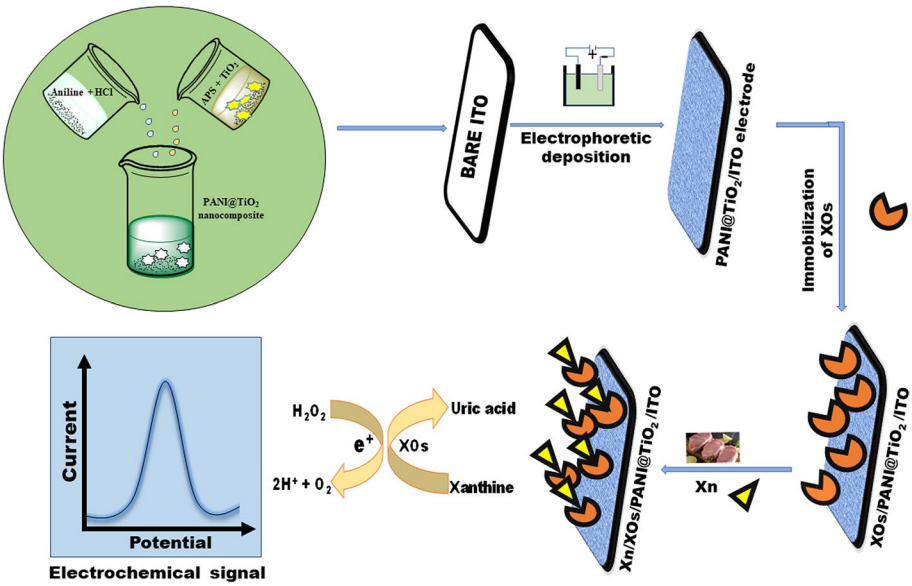


Fig. 1 Scheme showing the fabrication of PANI@TiO_2 nanohybrid–based biosensor

and Ag/AgCl as a reference electrode. All the electrochemical experiments were performed in phosphate buffer saline (PBS, 100 mM) pH 7.4 containing 5 mM $[Fe(CN)_6]^{3-/4-}$.

Results and Discussion

X-ray diffraction studies for PANI, TiO_2, and PANI@TiO_2 nanohybrid are shown in Fig. 2A. The broad reflection peak that appears around 20°–24° indicates that PANI was partially crystalline (curve (i)). The characteristic peaks at 25.5°, 38.14°, and 48.09° corresponded to the 101, 004, and 200 planes, respectively, confirming the presence of TiO_2 nanoparticles in PANI (curve (iii)). However, the pattern of the synthesized PANI@TiO_2 nanohybrid was found to be identical with TiO_2 (curve (ii)) and does not reflect any typical peak of PANI pattern. This may be due to the highly crystalline behavior of TiO_2 which overpower the intensity of the characteristic peaks of PANI [27].

The FT-IR spectra of PANI and PANI@TiO_2 nanohybrid displays benzenoid and quinonoid ring stretching bands at 1602 and 1482 cm^{-1}, respectively. The peaks at 2865 and 2853 cm^{-1} depict C–H stretching vibrations. Whereas, the bands at 1171, 1403, and 822 cm^{-1} are credited to C–N, C=N bond stretching, and C–H (benzene ring) out of the plane bending vibrations, respectively. The broad absorption bands (1000 to 550 cm^{-1}) is due to Ti–O–Ti vibration has been observed in PANI@TiO_2 nanohybrid, confirming the effective incorporation of TiO_2 into the PANI matrix (Fig. 2B (curve ii)) [28, 29].

The thermal stability of the nanohybrid has been studied using TGA. Figure 2C shows the mass losses of PANI and PANI@TiO_2 nanohybrids, heated at N_2 atmosphere. Initially,

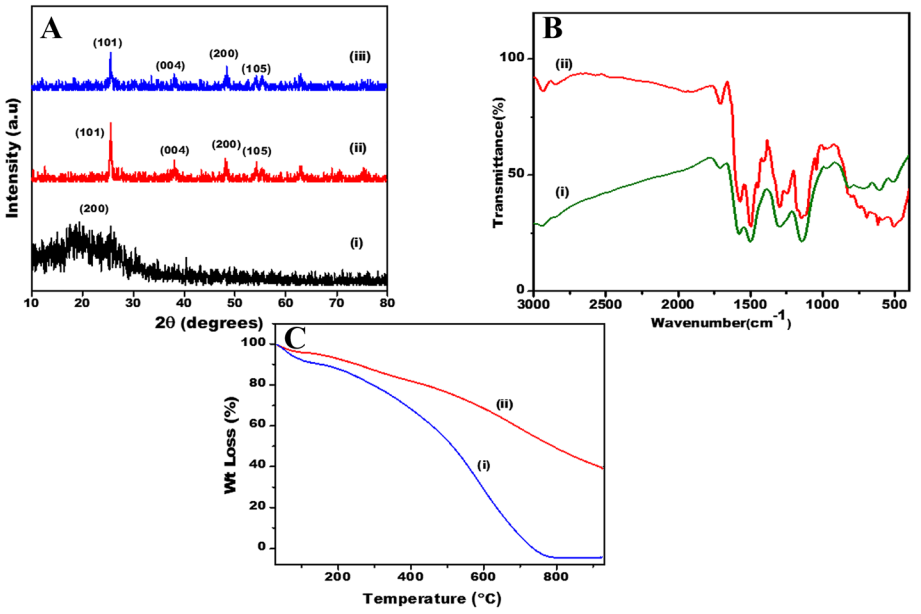


Fig. 2 **A** XRD of (i) PANI, (ii) TiO_2 nanoparticle, and (iii) PANI@TiO_2 nanohybrid; **B** FT-IR spectra of (i) PANI and (ii) PANI@TiO_2 nanohybrid; and **C** TGA curve for (i) PANI and (ii) PANI@TiO_2 nanohybrid in the range 25 °C–900 °C (10 °C/min), in a nitrogen environment

the loss in mass corresponds to the desorption of the moisture present in the material. PANI displays a rapid decrease in mass (10%) at a temperature of 280 °C due to the breakaway of dopant bonds (curve i). After that, a steady decrease in mass occurs in the range of 280 °C–600 °C. Whereas, the PANI@TiO_2 nanohybrid exhibits good stability with a slight loss in mass (4%) from 100 °C to 400 °C, and after 400 °C, the degradation rate increases (curve ii). The loss in mass of PANI may be due to the breakage of the polymer chain in the temperature range 280 °C to 600 °C. While the loss of mass in PANI@TiO_2 nanohybrid is slightly lower because TiO_2 restricts the flexibility of PANI chains, making them more stable than pristine PANI. The mass loss at 600 °C was found to be smaller for the PANI@TiO_2 nanohybrid, which indicates that the synthesized nanohybrid has good thermal behavior compared to the bare PANI [30].

The SEM image of PANI@TiO_2/ITO (Fig. 3A) shows that PANI@TiO_2 nanohybrid is evenly distributed on the surface of the ITO electrode [16]. After the immobilization of XOs on the PANI@TiO_2/ITO electrode, there was a structural change in the morphology, confirming that the enzyme has been evenly spread all over the PANI@TiO_2/ITO surface (Fig. 3B). Some globular-like structure observed in the SEM image is perhaps due to the agglomeration of XO on the surface of PANI@TiO_2/ITO electrode.

Electrochemical Studies of the Fabricated Electrodes

The electrochemical studies of bare ITO, PANI/ITO, and PANI@TiO_2/ITO electrodes have been performed using electrochemical impedance spectroscopy (EIS; PBS 5 mM of $[Fe(CN)_6]^{3-/4-}$) in the frequency range of 0.01–10^5 Hz [16]. In the Nyquist plot, a linear region is observed at lower frequencies and a semicircle region at higher frequencies, which corresponds to the diffusion-controlled electron transfer resistance process and electron transfer resistance–limited process, respectively [31]. The equivalent circuit (Fig. 4A (inset)) consists of solution resistance (R_s), an electron transfer resistance (R_{ct}), Warburg impedance (Z_w), and constant phase element (CPE) [18]. The R_{ct} of the ITO electrode was found to be 500.5 Ω (curve (i)), whereas the R_{ct} for PANI/ITO electrode and PANI@TiO_2/ITO electrode was found to be 123.9 Ω (curve (ii)) and 60.09 Ω (curve iii), respectively. The synergistic impact of PANI and TiO_2 provides a direct

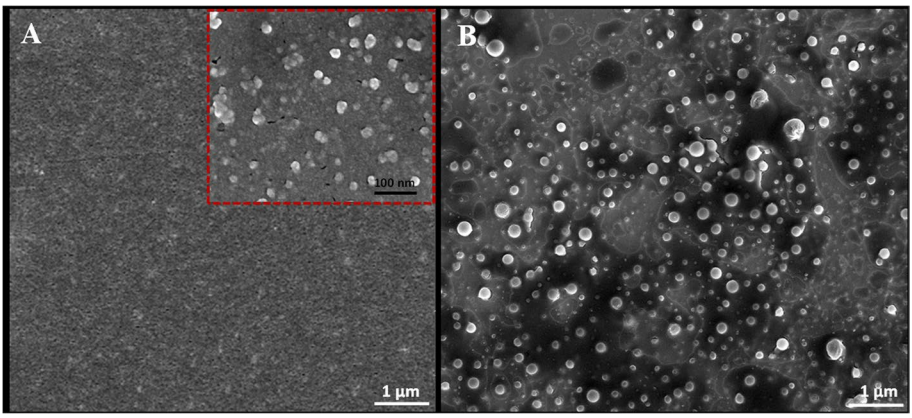


Fig. 3 SEM images of **A** PANI@TiO_2/ITO electrode and **B** XOs/PANI@TiO_2/ITO electrode

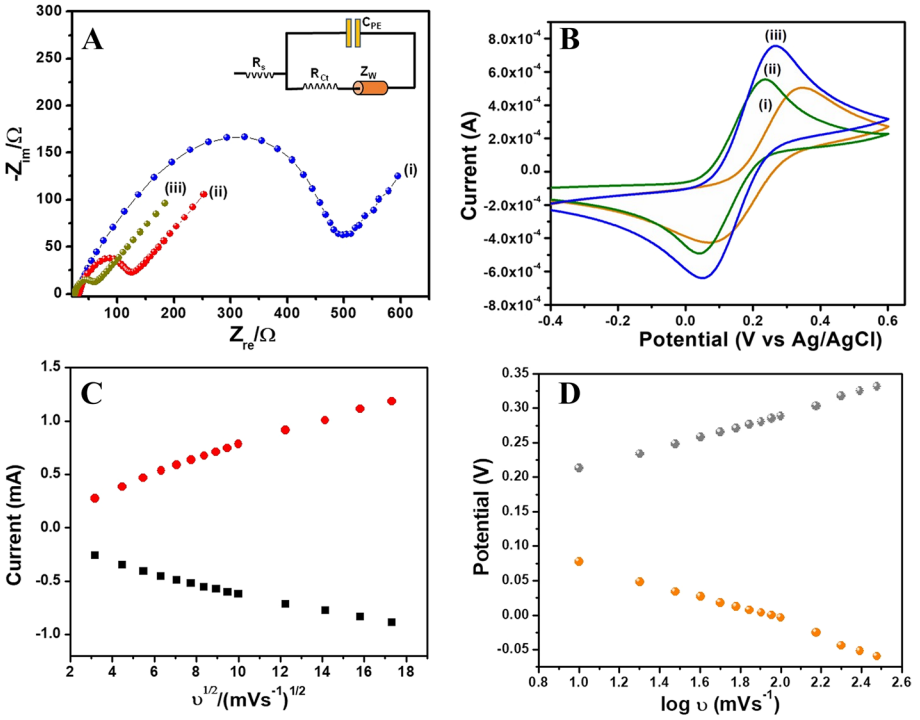


Fig. 4 **A** Nyquist diagram for the Faradic impedance for (i) ITO electrode, (ii) PANI/ITO electrode, and (iii) PANI@TiO_2/ITO electrode measured in PBS solution (pH 7) containing 5 mM $[Fe(CN)_6]^{3-/4-}$, in the frequency range from 10^5 to 10^{-2} Hz. **B** CV studies of (i) PANI/ITO electrode, (ii) PANI@TiO_2/ITO electrode, and (iii) XOs/PANI@TiO_2/ITO electrode. The plots of **C** I_{pa}, I_{pc} vs. square root of scan rate and **D** potential with log of scan rate with varying scan rate (10–300 mV) for PANI@TiO_2/ITO electrode

channel for charge transfer and a controlled diffusion process of the mediator ion, resulting in a lower R_{ct} value for the PANI@TiO_2/ITO electrode as compared to the PANI electrode [32].

From the obtained R_{ct} values, the electrochemical reaction that takes place at the electrode can be investigated by electron-transfer kinetics. The exchange current per geometric unit area (i_0) and apparent electron transfer rate constant (K_{app}) for PANI/ITO and PANI@TiO_2/ITO electrodes have been calculated using Eqs. (1) and (2).

$$i_0 = nRT/R_{ct}F \tag{1}$$

$$K_{app} = RT/n^2F^2AR_{ct}C \tag{2}$$

The i_0 for the PANI/ITO electrode was found to be 20.72×10^{-5} Acm^{-2}, whereas for the PANI@TiO_2/ITO electrode, i_0 calculated was 42.73×10^{-5} Acm^{-2}. Similarly, K_{app} for the PANI/ITO and PANI@TiO_2/ITO electrodes has been calculated to be 1.705×10^{-6} cm s^{-1} and 3.515×10^{-6} cm s^{-1}, respectively. This increase in values, i_0 and K_{app} of the PANI@TiO_2/ITO electrode compared to the PANI/ITO electrode, directly explains the precise diffusion of redox couple through the electrode interface [18].

Cyclic voltammetric studies have been carried out for PANI/ITO, PANI@TiO_2/ITO, and XOs/PANI@TiO_2/ITO electrodes in PBS buffer (Fig. 4B). It has been seen that the redox peak current of the PANI@TiO_2/ITO electrode (curve (ii); 0.49 mA) is higher compared to PANI/ITO electrode (curve (i)). It can be inferred that TiO_2 improves the conductivity of PANI@TiO_2/ITO electrode when compared to PANI/ITO electrode [33]. After the enzyme (XOs) is immobilized on PANI@TiO_2/ITO electrode, the redox peaks become more dominant (curve (iii); 2.02 mA) and the peak current value shifts to the positive side. This signifies that a direct electron pathway was taking place between the active site of XOs and the electrode, besides more enzymes are in interaction with the redox couple, thus indicating an increase in the oxidation signal [34].

The effect of changing the scan rate on the PANI@TiO_2/ITO electrode was explored, and it was observed that the rise in the scan rate from 10 to 300 mV/s resulted in a linear increase in oxidation peak current and a positive shift in the oxidation peak potential. Both the peak anodic (I_{pa}) and cathodic (I_{pc}) currents increases linearly with increasing scan rate and follow Equations (3) and (4), showing surface adsorption-controlled process kinetics (Fig. 4C).

$$Ipa(\mu A) = 1.2839 \times 10^{-4} + 6.3103 \times 10^{-5}\sqrt{v(mV/s)} \quad R = 0.9957 \tag{3}$$

$$Ipc(\mu A) = -1.7101 \times 10^{-4} - 4.2967 \times 10^{-5}\sqrt{v(mV/s)} \quad R = 0.9929 \tag{4}$$

Similarly, a linear correlation was observed for anodic and cathodic peak potentials (E_{pa} and E_{pc}) for increasing the natural logarithm of scan rate (ln$v$). The linear regression equations depicting the same have been given as Eqs. (5) and (6), respectively [35] (Fig. 4D).

$$Epa(V) = 0.1328\ln(v) + 0.0792; R = 0.9755 \tag{5}$$

$$Epc(V) = 0.1742ln(v) - 0.0928; R = 0.9868 \tag{6}$$

Using Laviron's model (Eq. 7), the charge transfer coefficient (α) and heterogeneous electron transfer coefficient (K_s) for PANI@TiO_2/ITO electrode have been calculated to be 0.9010 s^{-1} and 0.0772 s^{-1}, respectively, at a scan rate (v) of 50 mV/s.

$$ln(K_s) = \alpha ln(1-\alpha) + (1-\alpha)ln(\alpha) - ln(\frac{RT}{nFv}) - \alpha(1-\alpha)\frac{nF\Delta E_p}{RT} \tag{7}$$

Using linearity curves and Randles Sevick equation, Eq. (8), the diffusion coefficient (D) for the PANI@TiO_2/ITO electrode was found to be 5.48×10^{-11} cm^2 s^{-1}.

$$I_p = (2.99 \times 10^5)\alpha^{1/2}n^{3/2}ACD^{1/2}v^{1/2} \tag{8}$$

where C is the molar concentration of $[Fe(CN)_6]^{3-/4-}$ and A is the electrode area. The different kinetic parameters calculated for the PANI@TiO_2/ITO and XOs/PANI@TiO_2/ITO electrodes have been summarized in Table ST1. The observed improved electrochemical performance of the biosensor could be due to the improved conductivity of PANI@TiO_2 nanohybrid against bare PANI, which enhance the charge transfer rate and provides a single channel for interaction with the enzyme.

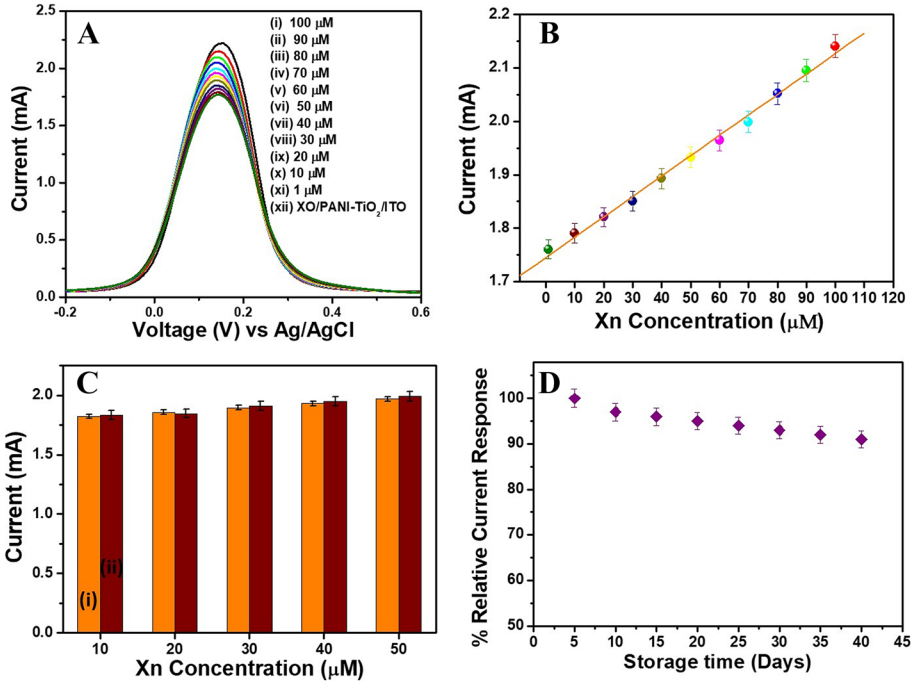


Fig. 5 **A** DPV studies showing the response of the XOs/PANI@TiO_2/ITO electrode with increasing concentration of Xn (from bottom to top, 1–100 µM); **B** calibration plot showing a linear relationship between the magnitudes of current recorded and concentration of Xn; **C** real sample analysis of biosensor with (i) buffer sample and (ii) real fish sample; and **D** stability studies of the fabricated biosensor

Electrochemical Biosensing Response Studies of XOs/PANI@TiO_2/ITO Electrode

The electrochemical biosensing studies of the XOs/PANI@TiO_2/ITO electrode have been carried out in the presence of different Xn concentrations (Fig. 5A). When the concentration of Xn was increased from 1 to 100 µM, the peak current response increased linearly. This increase in current indicates that there is a direct electron transfer between the Xn and the XOs/PANI@TiO_2/ITO electrode, which exhibits a good linear relationship between the Xn concentration and the current response (Fig. 5B). The sensitivity of the fabricated biosensor calculated using the regression equation: $y = 1.7704x + 0.4156$ (y = peak current and x = concentration of Xn) has been found to be 0.415 µA M^{-1} with $R^2 = 0.9942$. The limit of detection calculated using the Eq. $3\sigma/S$ was found to be 0.1 µM, where σ is standard deviation and S is the sensitivity obtained from the slope of the calibration curve [18].

Validation of the Biosensor with Real Sample

The practicality of the fabricated PANI@TiO_2/ITO biosensor for analyzing fish meat samples was conducted using DPV. The extraction of Xn from rohu (*Labeo rohita*) fish and the preparation of the Xn solution have been carried out using a simple method proposed by Watanabe et al [36]. In brief, fresh rohu fish was chopped and centrifuged at 4000 rpm in

Table 1 Comparison of the performance of PANI@TiO_2/ITO with other CP-based biosensors for Xn detection

S. No	Immobilization matrix	Detection technique	Linear range	LOD	Ref
1	Poly (L-aspartic acid)/MWCNT bio-nanocomposite	DPV	0.005–50.0 µM	3.5×10^{-4} µM	[38]
2	Polypyrrole films	Amperometry	4.0×10^{-4} M	1.0 µM	[39]
3	Poly (amino hydroxy naphthalene sulfonic acid)/GCE	LSV	4.0×10^{-4} –3.0×10^{-6} M	7.1×10^{-7} M	[40]
4	Polypyrrole-polyvinyl sulfonate film	Amperometry	10^{-7}–10^{-3} M	1.0×10^{-7} M	[36]
5	Nano-$CaCO_3$/GCE	CV	2×10^{-6}–2.5×10^{-4} M	2 µM	[41]
6	Ag-doped ZnO nanoparticles/polypyrrole	EIS Chronoamperometry	0.06–0.6 µM	0.07 µM	[42]
7	Graphene-TiO_2	CV	20–512 µM	9.5 µM	[43]
8	Poly-TTCA/Au	CV	5.0×10^{-6} – 1.0×10^{-4} M	1 µM	[44]
9	PANI@TiO_2/ITO	DPV	1–100 µM	0.1 µM	Present work

15-mL distilled water. Then the mixture was filtered using Whatman filter paper and was diluted 100 times to get the desired results. Electrochemical studies show that there was a continuous increase in the peak current with an increase in the Xn concentration from 10 to 50 μM and the results were similar to that of the synthetic Xn samples (Fig. 5C) [37]. The different biosensing parameters of the XOs/PANI@TiO_2-based biosensor were compared with previously reported polymer-based biosensors for Xn detection (Table 1), and it was observed that the fabricated biosensor shows an excellent response to Xn with low LOD and can detect the presence of Xn in fish meat sample. The proposed biosensor was used to detect the freshness of fish extract by evaluating the current response over a period of time span from 0 to 8 days at room temperature. As the storage duration of fish extract was extended, the level of xanthine increased, and it was almost six times higher on day 8 than it was on day 1 (Fig. S1).

Specificity, Reproducibility, and Stability Studies

The interference studies of the biosensor have been carried out by incubating the XOs/PANI@TiO_2 electrode with the same concentration of other interfering analytes (ascorbic acid, urea, and glucose). It was observed that there was no change in the current when the biosensor was subjected to other analytes revealing the specificity of the biosensor for Xn detection (Fig. S2). Further, the biosensor was also incubated with a mixture of the interfering analytes containing Xn. The electrochemical results indicate that the biosensor was able to detect the presence of Xn even in the presence of other interfering analytes. The reproducibility of the fabricated biosensor has also been measured by a series of repetitive voltammetric analyses. The % recoveries calculated for the XOs/PANI@TiO_2/ITO electrode ranged from 96 to 99%, with an RSD value of ≈ 4% ($n=4$), which was obtained for a 10-μM Xn measurement, demonstrating the fabricated biosensor shows acceptable reproducibility. The stability of the XOs/PANI@TiO_2/ITO electrode was tested by monitoring the DPV response of the bioelectrode ($n=3$) to Xn (10 μM) every 5 days while it was stored at 4 °C. After 45 days of use, the biosensor lost around 8% of its original activity, indicating that the biosensor is stable for at least 40 days when stored at 4 °C in a dry state (Fig. 5D).

Conclusion

An electrochemical enzymatic biosensor has been developed for Xn detection using PANI@TiO_2 nanohybrid. The fabricated biosensor shows improved electrochemical properties, a reasonable detection limit (0.1 μM), and quick response time (10 s). The validation of the biosensor with real sample obtained from rohu fish shows its potential for field applications for detection of Xn in fish meat. The fabricated biosensor was found to be specific, stable (40 days), and reproducible. The developed PANI@TiO_2 nanohybrid–based platform has been successfully validated with the fish meat sample, which shows its applicability and reliability for fish freshness detection.

Abbreviations MWCNT: Multiwalled carbon nanotubes; XO: Xanthine oxidase; $CaCO_3$: Calcium carbonate; GCE: Glassy carbon electrode; Ag: Silver; ZnO: Zinc oxide; poly-TTCA: Poly-5,2′:5′,2″-terthiophine-3-carboxylic acid; Au: Gold; TiO_2: Titanium dioxide; ITO: Indium tin oxide; DPV: Differential pulse voltammetry; CV: Cyclic voltammetry; EIS: Electrochemical impedance spectroscopy

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s12010-022-03931-7.

Acknowledgements Deeksha acknowledges DST-INSPIRE for the SRF Award. C.M. Pandey acknowledges the Department of Science and Technology (DST), New Delhi, India for the DST-INSPIRE Faculty award.

Author Contribution Deeksha: conceptualization, methodology, data curation, formal analysis, writing—original draft. CMP: supervision, investigation, validation, writing—review and editing, funding acquisition. DK: supervision, validation, project administration, writing—review and editing.

Data Availability Not applicable.

Declarations

Ethics Approval Not applicable.

Consent to Participate Not applicable.

Consent for Publication Not applicable.

Competing Interests The authors declare no competing interests.

References

1. Kamil Reza, K., Singh, M. K., Yadav, S. K., Singh, J., Agrawal, V. V., & Malhotra, B. D. (2013). Quantum dots based platform for application to fish freshness biosensor. *Sensors and Actuators B: Chemical, 177*, 627–633.
2. Sahyar, B. Y., Kaplan, M., Ozsoz, M., Celik, E., & Otles, S. (2019). Electrochemical xanthine detection by enzymatic method based on Ag doped ZnO nanoparticles by using polypyrrole. *Bioelectrochemistry, 130*, 107327.
3. Devi, R., Yadav, S., & Pundir, C. S. (2011). Electrochemical detection of xanthine in fish meat by xanthine oxidase immobilized on carboxylated multiwalled carbon nanotubes/polyaniline composite film. *Biochemical Engineering Journal, 58–59*, 148–153.
4. Albelda, J. A. V., Uzunoglu, A., Santos, G. N. C., & Stanciu, L. A. (2017). Graphene-titanium dioxide nanocomposite based hypoxanthine sensor for assessment of meat freshness. *Biosensors and Bioelectronics, 89*, 518–524.
5. Lawal, A. T., & Adeloju, S. B. (2009). Comparison of enzyme immobilisation methods for potentiometric phosphate biosensors. *Biosensors and Bioelectronics, 25*, 406–410.
6. Ghanbari, K., & Nejabati, F. (2020). Ternary nanocomposite-based reduced graphene oxide/chitosan/Cr2O3 for the simultaneous determination of dopamine, uric acid, xanthine, and hypoxanthine in fish meat. *Analytical Methods, 12*, 1650–1661.
7. Jalil, O., Pandey, C. M., & Kumar, D. (2020). Electrochemical biosensor for the epithelial cancer biomarker EpCAM based on reduced graphene oxide modified with nanostructured titanium dioxide. *Microchimica Acta, 187*, 275.
8. Yadav, S. K., Singh, J., Agrawal, V. V., & Malhotra, B. D. (2012). Nanostructured nickel oxide film for application to fish freshness biosensor. *Applied Physics Letters, 101*, 023703.
9. Solanki, S., Pandey, C. M., Gupta, R. K., & Malhotra, B. D. (2020). Emerging trends in microfluidics based devices. *Biotechnology Journal, 15*, 1900279.
10. Xue, G., Yu, W., Yutong, L., Qiang, Z., Xiuying, L., Yiwei, T., & Jianrong, L. (2019). Construction of a novel xanthine biosensor using zinc oxide (ZnO) and the biotemplate method for detection of fish freshness. *Analytical Methods, 11*, 1021–1026.
11. Gerard, M., Chaubey, A., & Malhotra, B. D. (2002). Application of conducting polymers to biosensors. *Biosensors & Bioelectronics, 17*, 345–359.
12. Kumar, A. S., & Shanmugam, R. (2011). Simple method for simultaneous detection of uric acid, xanthine and hypoxanthine in fish samples using a glassy carbon electrode modified with as commercially received multiwalled carbon nanotubes. *Analytical Methods, 3*, 2088–2094.

13. Pundir, C. S., & Devi, R. (2014). Biosensing methods for xanthine determination: A review. *Enzyme and Microbial Technology, 57*, 55–62.
14. Dhand, C., Das, M., Datta, M., & Malhotra, B. D. (2011). Recent advances in polyaniline based biosensors. *Biosensors & Bioelectronics, 26*, 2811–2821.
15. Ahuja, T., Mir, I. A., Kumar, D., & Rajesh. (2007). Biomolecular immobilization on conducting polymers for biosensing applications. *Biomaterials, 28*, 791–805.
16. Soni, A., Pandey, C. M., Pandey, M. K., & Sumana, G. (2019). Highly efficient polyaniline-MoS2 hybrid nanostructures based biosensor for cancer biomarker detection. *Analytica Chimica Acta, 1055*, 26–35.
17. Solanki, P., Kaushik, A., Agrawal, V., & Malhotra, B. (2011). Nanostructured metal oxide-based biosensors. *NPG Asia Materials, 3*, 17–24.
18. Jalil, O., Pandey, C. M., & Kumar, D. (2021). Highly sensitive electrochemical detection of cancer biomarker based on anti-EpCAM conjugated molybdenum disulfide grafted reduced graphene oxide nanohybrid. *Bioelectrochemistry, 138*, 107733.
19. Verma, R., Samdarshi, S. K., Sagar, K., & Konwar, B. K. (2017). Nanostructured bi-phasic TiO2 nanoparticles grown on reduced graphene oxide with high visible light photocatalytic detoxification. *Materials Chemistry and Physics, 186*, 202–211.
20. Teli, S. B., Molina, S., Sotto, A., Calvo, E. G., & Abajob, J. D. (2013). Fouling resistant polysulfone–PANI/TiO2 ultrafiltration nanocomposite membranes. *Industrial & Engineering Chemistry Research, 52*, 9470–9479.
21. Xie, S., Gan, M., Ma, L., Li, Z., Yan, J., Yin, H., Shen, X., Xu, F., Zheng, J., Zhang, J., & Hu, J. (2014). Synthesis of polyaniline-titania nanotube arrays hybrid composite via self-assembling and graft polymerization for supercapacitor application. *Electrochimica Acta, 120*, 408–415.
22. Soni, A., Pandey, C. M., Solanki, S., Kotnala, R. K., & Sumana, G. (2018). Electrochemical genosensor based on template assisted synthesized polyaniline nanotubes for chronic myelogenous leukemia detection. *Talanta, 187*, 379–389.
23. Katoch, A., Burkhart, M., Hwang, T., & Kim, S. S. (2012). Synthesis of polyaniline/TiO2 hybrid nanoplates via a sol–gel chemical method. *Chemical Engineering Journal, 192*, 262–268.
24. Rab, N., Chong, F. K., Mohamed, H. I., & Lim, W. H. (2018). Preparation of TiO2 nanoparticles by hydrolysis of TiCl4 using water and glycerol solvent system. *Journal of Physics: Conference Series, 1123*, 012065.
25. Paul, G., Verma, S., Jalil, O., Thakur, D., Pandey, C. M., & Kumar, D. (2021). PEDOT: PSS-grafted graphene oxide-titanium dioxide nanohybrid-based conducting paper for glucose detection. *Polymers for Advanced Technologies, 32*, 1774–1782.
26. Elakkiya, S., Arthanareeswaran, G., Ismail, A. F., Das, D. B., & Suganya, R. (2019). Polyaniline coated sulfonated TiO2 nanoparticles for effective application in proton conductive polymer membrane fuel cell. *European Polymer Journal, 112*, 696–703.
27. Yu, H., Jang, K., Chung, I., & Ahn, H. (2016). Fabrication and electrochemical characterization of polyaniline/titanium oxide nanoweb composite electrode for supercapacitor application. *Journal of Nanoscience and Nanotechnology, 16*, 2937–2943.
28. Najafi, V., Ahmadi, E., Ziaee, F., Omidian, H., & Sedaghat, H. (2019). Polyaniline-modified TiO2, a highly effective photo-catalyst for solid-phase photocatalytic degradation of PVC. *Journal of Polymers and the Environment, 27*, 784–793.
29. Ahmad, R., & Mondal, P. K. (2012). Adsorption and photodegradation of methylene blue by esing PANI/TiO2 nanocomposite. *Journal of Dispersion Science and Technology, 33*, 380–386.
30. Emran, K. (2018). The electrocatalytic activity of polyaniline/TiO2 nanocomposite for Congo red degradation in Aqueous solutions. *International Journal of Electrochemical Science, 13*, 5085–5095.
31. Abaci, S., Nessark, B., & Riahi, F. (2014). Preparation and characterization of polyaniline+TiO2 composite films. *Ionics, 20*, 1693–1702.
32. Reddy, R., kv, K., S B, B., Soni, S., Jeong, H. M., and Anjanapura, R. (2016). Enhanced photocatalytic activity of nanostructured titanium dioxide/polyaniline hybrid photocatalysts. *Polyhedron, 120*, 169–174.
33. Majumdar, S., & Mahanta, D. (2020). Deposition of an ultra-thin polyaniline coating on a TiO2 surface by vapor phase polymerization for electrochemical glucose sensing and photocatalytic degradation. *RSC Advances, 10*, 17387–17395.
34. Elgrishi, N., Rountree, K. J., McCarthy, B. D., Rountree, E. S., Eisenhart, T. T., & Dempsey, J. L. (2018). A Practical beginner's guide to cyclic voltammetry. *Journal of Chemical Education, 95*, 197–206.

35. Pandey, C. M., Sumana, G., & Tiwari, I. (2014). Copper oxide assisted cysteine hierarchical structures for immunosensor application. *Applied Physics Letters, 105*, 103706.
36. Dolmacı, N., Çete, S., Arslan, F., & Yaşar, A. (2012). An amperometric biosensor for fish freshness detection from xanthine oxidase immobilized in polypyrrole-polyvinylsulphonate film. *Artificial Cells, Blood Substitutes, and Biotechnology, 40*, 275–279.
37. Narang, J., Malhotra, N., Singhal, C., Singh, M., & Pundir, C. (2016). Impedimetric and voltammetry sensing of xanthine using nanocomposites. *Advanced Materials Letters, 7*, 555–560.
38. Yazdanparast, S., Benvidi, A., Abbasi, S., & Rezaeinasab, M. (2019). Enzyme-based ultrasensitive electrochemical biosensor using poly(l-aspartic acid)/MWCNT bio-nanocomposite for xanthine detection: A meat freshness marker. *Microchemical Journal, 149*, 104000.
39. Arslan, F., Yaşar, A., & Kiliç, E. (2006). An amperometric biosensor for xanthine determination prepared from xanthine oxidase immobilized in polypyrrole film. *Artificial Cells, Blood Substitutes, and Immobilization Biotechnology, 34*, 111–126.
40. Mathew, M. R., & Girish Kumar, K. (2020). Poly(amino hydroxy naphthalene sulphonic acid) modified glassy carbon electrode; an effective sensing platform for the simultaneous determination of xanthine and hypoxanthine. *Journal of The Electrochemical Society, 167*, 047519.
41. Shan, D., Wang, Y., Xue, H., & Cosnier, S. (2009). Sensitive and selective xanthine amperometric sensors based on calcium carbonate nanoparticles. *Sensors and Actuators B: Chemical, 136*, 510–515.
42. Sahyar, B., Kaplan, M., Ozsoz, M., Celik, E., & Otles, S. (2019). Electrochemical xanthine detection by enzymatic method based on Ag doped ZnO nanoparticles by using polypyrrole. *Bioelectrochemistry, 130*, 107327.
43. Albelda, J. A. V., Uzunoglu, A., Santos, G. N. C., & Stanciu, L. A. (2017). Graphene-titanium dioxide nanocomposite based hypoxanthine sensor for assessment of meat freshness. *Biosensors and Bioelectronics, 89*, 518–524.
44. Rahman, M. A., Won, M.-S., & Shim, Y.-B. (2007). Xanthine sensors based on anodic and cathodic detection of enzymatically generated hydrogen peroxide. *Electroanalysis, 19*, 631–637.

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

International Journal of Multidisciplinary Research and Growth Evaluation.

Hydro-Electric Power Generation

Mukul Chauhan [1], Neeraj Joshi [2], JP Kesari [3]
[1] B.Tech, Dept. of Software Engineering, Delhi Technological University, Delhi, India
[2] B.Tech, Dept. of Software Engineering, Delhi Technological University, Delhi, India
[3] Faculty of Mechanical Engineering, Delhi Technological University, Delhi, India

* Corresponding Author: **Mukul Chauhan**

Article Info

ISSN (online): 2582-7138
Volume: 03
Issue: 02
March-April 2022
Received: 23-03-2022;
Accepted: 12-04-2022
Page No: 544-546

Abstract
In the present period where we are confronting an emergency and delayed consequences of overconsumption of traditional fuel sources, it is the need of great importance to change to green and clean choices of energy like hydro-electric power energy and so forth which don't harm the climate by their harmful discharges of poisonous oxygen-compounds. Hydro-Electric power is that the energy extracted from running water-streams, waterways. It is one in everything about the most seasoned techniques for energy creation. Indeed, even during primitive times, individuals used to get energy from water wheels. The commitment of hydel power on the planet's energy-creation situation is tremendous expanding. Yet, the business creation of hydro-electric power requires an enormous framework because of which a great deal of grief that is caused to the biological system, Gravitational Water Vortex Power Plant (G.W.V.P.P.) is presented which lessens the viable expense by diminishing foundation.

Keywords: Hydro-electric power, Notch, Primitive, Whirlpool

1. Introduction
A gravitational water whirlpool power plant is a strategy for the age of hydro-electric power that utilizes a straightforward system. In this plant, water is brought into a basin that is circular in shape, which makes a free whirlpool and energy is being removed from the free whirlpool by utilizing a turbine. The advantages of utilizing this falsely evolved vortex over this gravity-sped up water help to increase the proficiency, decrement in cost, and not just the adverse consequence on the climate is decreased, however, helps in expanding the manageability and state of the waterway overall. Whirlpool is a design that helps in the arrangement of a gravitational whirlpool stream.



Fig 1

2. Component Review inlet and outlet Structure

The outlet cannal waterway is normally located in the bowl's centre, and the size of the outlet channel has a considerable impact on the vortex turbine's strength and effectiveness. Extraneously, the intake stream trench conducts the water stream into the bowl. In general, an inlet has two main considerations: the first is the height of the water channel, and the second is the channel width, which is used to determine the mass stream rate of water for energy generation.

Turbine

It is the even possible that the main part of G.W.V.P.P. which uses the dynamic energy of water emerging from the basin to change the active energy over to electric energy.it is typically positioned at the center of basin simply over the watercourse that serves as an outlet source Because the concept of GWVPP isn't well-founded, different analysts use various types of vortex turbine plants for their proffer assignment. According to recent research, directed most extreme productivity is obtained when the turbine is placed close to the power source trench. When we vary the turbine location and cutting edge math, which is still a high level subject of examination in this industry, we notice an overall variation of 14-25 percent.

Configuration of the basin

The basin's plan determines the vortex profile that will be created during the cycle. As a result of many investigations, there are fundamentally three basin layouts that are widely recognised as -

1. Basin with a conical shaped
2. Basin with a cylindrical shape
3. Basin with rectangular shapes

According to various studies, a barrelshaped bowl is the most effective bowl plan for the G.W.V.P.P. because it produces a more uniform stream rate. When the speed of the vortex is estimated by varying the indent point.

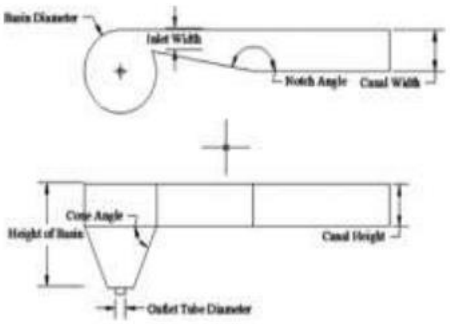


Fig 2

Working

The water is pumped into a tank and then into a flow tank. This combination of limited low-pressing factors at the because of the high speed and the ensuing commotion. Dissemination in the distracting section leads to the development of a strong whirlpool stream.The transition from anticipated energy to rotational energy. The whirlpool centre is a source of dynamic energy. An upward hub turbine then uses this. used The change of expected energy to rotational dynamic energy happens at the whirlpool center. This is then by an upward hub turbine.



Fig 3

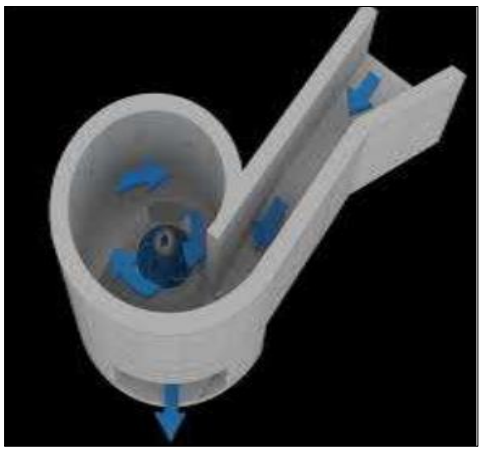


Fig 4

Advantages

This hydro-electric plant design has a few advantages over the traditional dams in use today:

1. Water circulation is normal due to the high speed of the water stream and the increase in the water surface area, which aids in the self-sanitization of water and the enhancement of it with basic microorganisms and water plants.
2. It is more practicable and feasible to develop on a low-budget basis than the traditional time dam with its massive infrastructure requirements.
3. It does not disrupt the oceanic daily routine or the local community's experience of being close to a water supply.
4. Aids in the decomposition of oxygen.
5. Because this venture may be started at a small scale and does not require a large amount of space, the inventory can be guaranteed.

Mathematical formulation

P=Hv * Q * d * g Watt(W) ,
g = gravitational acceleration
d = Density of water(kg/m3)
P = Power Hv = Height of vortex

Setting UP the project

There is a 5.4KW SUEZ WATER TREATMENT PLANT in Versailles, France a facility that has been added to the wastewater with an apparent treatment plant at Versailles 0.65m3 /s progression, as well as a water pressure of 3.1m.The turbine is set up to provide electricity to the house. A compound-free wastewater called Carré de Réunion. Versailles, France's wastewater treatment plant.



Fig 5

CUNCO, CHILE

In Cunco, Chile, a model of this model was built. This turbine had a 1.7-meter head and a 1.8-meter-per-second progression.

Stream rate and absolute effectiveness factor determine the electric force of a hydroelectric power plant. A hydropower plant is typically designed to operate for 5000 full burden hours per year. As a result, the electric force and the same full burden hours can be used to determine the annual production limit.The turbine's productivity was found to be 80%.In its first year of operation, the turbine produced 55 kWh.

Green School

Bali, Indonesia-The GWVPP power plant has been arranged on the planet renowned green school situated close to Ayung stream in Bali, Indonesia. The 14 KW vortex turbine arrangement advantages in excess of 750 undergrads and staff. This task simply requires 1.57m3 /s of water stream and vortex tallness of 1.85 m

4. Conclusion

The GWVPP is an expense proficient miniature hydro-electric power turbine used to create nearby inexhaustible and solid energy with no antagonistic effect on the weather. The turbine is structured uniquely for effective utilization of a huge stream with a few pressing factors considering greater resistances in the complete development of task and accomplishing all fish well-disposed boundaries. It is an easy-tounderstand the framework that can be utilized without any need of ability and since most segments are pre-manufactured, can be arranged with neighborhood work. Additionally, it is the need of great importance to begin utilization of such eco amicable fuel hotspots for the manageable turn of events.

Acknowledgement

The author wishes to express sincere gratitude to Delhi Technological University (DTU) and Dr. Janardan Prasad Kesari for their support and providing the data for this publication.

5. References

1. VK Gorshkosv. Knowledge of Environment. UNESCO Encyclopaedia of Life Support Systems, 2008.
2. Alternative Energy, “Hydropower originates from the sun,” Available from:http://www.alternativeenergyprimer.co m/Hydropower.html
3. W Luvtz, W Sanedersen, M Baeker, J Careter, C Wood. Earthiscan, London, UK, 1st edition.
4. M Toliba. The World Environment: Two Decades of Challenge, Chapman and Hall for UNEP, Nairobi, Kenya, 1972-1992.
5. Norwegiean Ministry of Petroleum and Energy. Facts 2008: energy and water resources in Norway, Norwegian Government Report. Available from:https://inis.iaea.org/search/search.aspx ?orig_q=RN:38001775
6. Koauris Centri. Turbine [cited 22nd September 2016]. Available from: http://www.kcthydropower.com

Image Compressed Sensing Using Deep Learning

Saksham Gera[1], Shiva Gupta[2], Paras HS[3], Rohan Pillai[4], Madan Mohan Rayguru[5]

Electrical Engineering Department,
Delhi Technological University
Shahbad Daulatpur Village, Rohini, New Delhi,
Delhi 110042, India

[1]sakshamgr3@gmail.com
[2]Shiva gupta email
[3]Paras HS email
[4]rohan92@dtu.ac.in
[5]mmrayguru@dtu.ac.in

***Abstract*—In this paper, techniques used to compress data, especially images are discussed. A deep learning model based on convolution neural networks is implemented and the results in the form of PSNR (Peak Signal to Noise Ratio) are compared. The main hurdles in compressed sensing of images are the sampling matrix design and the reconstruction method. To overcome these we propose a CNN framework called CS Net, which consists of a joined sampling network and reconstruction network. The sampling network adapts the sampling matrix during the training of data, which enables the compressed sensing measurements to maintain additional structural data for better recreation of images. The reconstruction network is divided into two parts, an elementary linear network, and a deep reconstruction non-linear network; it adapts a mapping between the endpoints of structural measurements and reconstructed images. In this project, we used a different number of blocks in the deep reconstruction and compared the results. The PSNR values of the reconstructed images gradually increase as we increase the number of residual blocks, but after a certain point, the values start decreasing[1].**

The results obtained show that CSNet produces high-quality images like those obtained from methods like FFT and Nyquist Shanon but with comparatively low computation costs[2][3].

***Keywords*— Compressed Sensing, Convolution Neural Networks, CSNet, Sampling matrix, Image reconstruction.**

I. INTRODUCTION

Image compression is a signal compression technique in which the real picture is reproduced using a limited amount of memory. Picture compaction aims to reduce image duplication and make data storage and transit extra affordable. The fundamental purpose of such a system is to decrease the maximum quantity of data that may be stored so that the decoded image is as close to the original as possible.

A. Overview

The classical image capturing system typically obtains a high volume of samples focused on the Nyquist-Shannon sampling theorem, with a sampling ratio of at least two times the signal's bandwidth, and then compresses the signal for storage and retrieval using a computationally complex compression method to remove redundancy. However, in certain image processing applications where data gathering devices must be simple, this type of image capture system may not be preferred.

Compressed Sensing (CS) is an emerging technique that illustrates a paradigm shift for picture storage and recovery that incorporates skimming and contraction. When a signal is sporadic in a domain, the compressed sensing theory indicates that a signal can be retrieved with many lesser observations than the Nyquist-Shannon sampling theorem suggests. It is commonly known that images contain a lot of unnecessary data and may be expressed sparsely. As a result, according to CS theory, an image may be compressed and rebuilt effectively.

Since the core CS study indicated that when a signal is sparse in a particular domain, it may be recreated with excellent clarity at a lower sampling ratio, there's been a great deal of interest in CS for picture capture. Image acquisition devices based on CS have been developed. The so-called single-pixel camera is the most well-known CS gadget among them. Other steps have been taken to improve the imaging versatility of CS-based cameras and to investigate their potential use in cellular devices and other portable devices.

The creation of the sample matrix and the design of the reconstruction method are the two primary issues in the research of CS. In the literature, several solutions for dealing with these two issues have been offered. To address the very first difficulty, many

sample matrices, such as the random matrix, binary matrix, and structural matrix, have been created. Furthermore, all of these matrices are input signal independent sampling matrices, meaning they disregard the signal's properties. Many sparsity-regularized-based solutions have been developed for the second issue, such as convex-optimization algorithms, greedy algorithms, and iterative thresholding algorithms. Certain efforts in image CS explored picture priors to create more complicated models; others, on the other hand, included more treatment efficiency in the repeated process of compression.

In our research paper, new variations on the CS framework based on convolution neural networks(namely CSNet) are tried and the results have been compared to find the most optimum number of residual blocks in the advanced reconstruction network.

We have used a Convolutional Neural Network (CSNet) to create a CS framework that contains a sampling network & a reconstruction network that are optimized together. The floating-point sampling matrix is foremost learned from the training pictures by the sampling network, allowing the CS measurements to preserve further image contextual data for improved recovery.

Among the readings of CS and the rebuilt pictures, the reconstruction network adapts an end-to-end point mapping. It comprises a remnant training-based non-linear deep reconstruction structure and an initial reconstruction network. Inter-block information may be successfully used by the reconstruction network, and blocking artifacts can be avoided.

Unlike classic CS approaches, the learned model does not require the sampling matrix to be transferred from encoder to decoder. Traditional image CS reconstruction algorithms can also benefit from the learned sampling matrix.

B. Principal of Image Compression

The fact that nearby pixels are connected, therefore, they contain duplicate information. It is a frequent feature of most images. The first and foremost step is to find a representation of the image that is less coupled. Redundancy and the reduction of irrelevant features are two key components of compression. Irrelevancy reduction removes sections of the signal that the receiver will not notice, such as the Human Visual System (HVS).

1) *Coding Redundancy*: The representation of information is linked to coding redundancy. Source codes serve as a representation of the information. When an image's grey levels are processed in a way that employs more coded symbols to represent each grey level than what is strictly necessary, the resulting picture is said to have coding redundancy.

2) *Spatial/Temporal Redundancy:* As maximum 2-D intensity array pixels are connected spatially; content is duplicated in the representations of the associated pixels. Sequentially connected pixels in a video clip also replicate information.

3) *Irrelevant Information:* Many 2-D intensity arrays carry data that the human supervision ignores and is irrelevant to the image's intended use. In the perspective that it isn't used, it is redundant.

Image compression research tries to reduce the bits needed to represent an image as much as possible by eliminating spatial and spectral redundancy.

C. Motivation

In recent years, the transmission and storage of data have increased by many folds. Data compressions techniques are used to reduce the storage size of data and allow faster transmission of larger data over the internet. Images are one of the most popular data types. Image compression deals with reducing the size of an image by eliminating redundant and unnecessary signals from its sparse matrix. State-of-the-art methods like Nyquist-Shannon or Fast Fourier Transform allows high-quality reconstruction of images but at the cost of computational power. Higher computational power is not available throughout and also slows the speed of the process. Therefore, the compressed sensing technique offers compression of data below Nyquist rates making it an attractive solution in facial recognition systems and medical imaging. Compressed sensing produces images of similar quality but at lesser computation time. Hence, the need for a proper deep learning model for compressed sensing.

II. BACKGROUND

In classical picture acquisition, the analog signal is captured by collecting a large number of data samples based on the Nyquist-Shannon sampling theorem.

Following the acquisition, the image (dense image) is converted over a basis Ψ (e.g., Fast Fourier transform of wavelet basis) with tiny coefficients for the majority of the places in the spatial domain of the image.

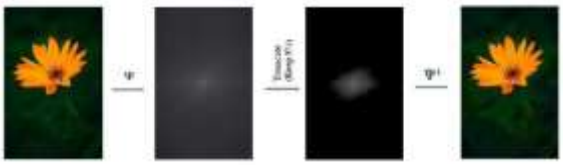


Figure 1An example showing traditional image compression and decompression using a basis function Ψ over a high-resolution image.

The pixels in the modified image are then shortened, removing a significant amount of data (95% in Fig. 1). The obtained picture (sparse image representation) may be easily stored as a dictionary of key-value pairs, with the key being the pixel's spatial orientation and the value being the transform coefficient. Compression is the term for this procedure.

Image data in the form of a dictionary is very straightforward to send over computer networks and consumes a smaller amount of storage.

The compressed sparse image is subjected to the inverse basis transform in order to rebuild the image. Decompression is the term for this procedure.

A. Compressed Sensing Background

Compressed sensing is a type of data collection that allows us to rebuild high-resolution pictures by sampling a significantly less amount of data than the usual image capture approach.

When a signal is sparse in some domains, the CS theory allows flat representation of a wide-dimensional signal into a dimension considerably tiny in contrast to the raw signal while permitting retrieval of the data from projections with precision.

Particularly, visualize that $x \in R^{N\times1}$ is a real-valued signal (original image) and $\phi \in R^{M\times N}$ is a sampling matrix. ($M << N$). The CS measurement collection process is written as

$$y = \phi x \quad (1)$$

where $y \in R^{M\times1}$ is the Compressed sensed metric (CS measurement). The CS theory indicates that successfully recuperating x is achievable because the signal x is scarce in a certain domain ψ.

The simplest way to describe CS reconstruction is to write it like this:

$$\min \|\psi_x\|p \text{ , s.t. } y = \phi x \quad (2)$$

The suffix *P* is normally assigned to 1 or 0, describing the sparseness of the vector ψ_x, where x are the sparing factors w.r.t to scope. In previous literature, a wide variety of techniques for tackling this optimization issue have been offered. The convex optimization approach, for example, converts a nonconvex issue into a convex one to obtain an estimated solution. The most frequent convex optimization approach for CS restoration is basis pursuit. It solves a linear programming issue by replacing the L_0 norm constraint with the L_1 norm constraint. Such convex-programming approaches, on the other hand, have a large computational cost.

Convex optimization, gradient-descent techniques, greedy algorithms, and Land weber (PL) algorithms are some of the methodologies presented in the literature for tackling this optimization issue. Despite their efforts, classical approaches have a high computational cost due to the necessity of repetitive computing[9]. Reconstructing a high-quality image using these CS approaches might take anything from a few seconds to many minutes. In addition, the sampling matrix from the encoder to the decoder should be sent



Figure 2An example showing recovering original image X by sampling a few pixels (Φx) while sensing and solving for the sparse representation over a basis Ψ.

A. Sparsity

The term sparsity, as the name implies, relates to the smallest quantity of data we can obtain and retain in order to reconstruct an original signal/document from it.

A sparse matrix is one with a large number of zero values. Sparse matrices are distinguished from dense matrices, which have a large number of non-zero values.

The fascination with sparsity stems from the fact that its use can result in significant computational savings, as well as the fact that many big matrix problems encountered in practice are sparse.

Sparsity should not be mistaken with missing data, because missing data indicates that a large number of data points are unknown. If the data is sparse, on the other hand, all of the data points are known, but the majority of them have little value[10].

A sparse matrix is a two-dimensional array of pixels with some pixels having zero intensity in terms of picture reduction. Sparsity is a technical term that refers to the percentage of cells in a database table that is empty.

Finally, all image compression techniques and methodologies aim to discover the optimum sparse matrix that can identify a particular image and compress it in the most optimal spatial and time complexity way possible.

B. Nyquist-Shannon Theorem

The Nyquist-Shannon theorem explains the sampling of a signal or waveform so that no information is lost[1][2].

Let's say we have a signal *X(t)*. On taking Fourier Transform of this signal, $X^{\Lambda}(f) = F\{x(t)\}$, there will be a f_{max} for which,

$$x|f| = 0 \forall |f| > f_{max} \quad (3)$$

such that beyond the maximum frequency f_{max} , there is no power in the signal. The Nyquist-Shannon theorem indicates that if we're to sample this signal, we'd require samples with a frequency greater than double the signal's peak frequency, that is,

$$f_{sample} > 2f_{max} \quad (4)$$

If that's the case, no information was sacrificed during the sampling procedure, and the source signal might be restored from the sampled signal.

This theorem is a stepping stone toward more advanced signal (image) processing, and it serves as a benchmark for various techniques and methodologies aimed at achieving the best image compression.

It also lays the groundwork for image compression algorithms that use a sampling frequency lower than Nyquist's.

III. PROPOSED WORK

The formulation of the sampling matrix and the design of reconstruction algorithms are the two key problems in the research of compressed sensing. In the literature, several solutions for dealing with these two issues have been offered. To address the first difficulty, many sampling matrices have been devised, including the random matrix, the binary matrix, and the structural matrix. However, each one of these sampling matrices are independent of the signal, meaning they disregard the signal's properties.

Many sparsity regularized-based approaches, such as convex optimization algorithms, greedy algorithms, and iterative thresholding algorithms, have been presented in response to the second difficulty. Several image CS works investigated pictures, prior to constructing more intricate models, while others added additional optimal factors into the recurrent thresholding procedure[4]. Some present state-of-the-art CS reconstruction methods that take numerous minutes to recreate a solitary slightly elevated picture, whilst others have much worse reconstruction accuracy. Deep learning has lately been shown to be more effective in computer vision tasks. There are indeed some deep learning-based picture CS reconstruction techniques that employ a random floating-point sampling matrix that we are aware of. However, because block-by-block reconstruction approaches only employ Intra block information to recreate a block, blocking artifacts emerge, and postprocessing is frequently required. Because of the usage of an orthogonal matching pursuit algorithm, the postprocessing approach has are relatively high computational cost[7].

IV. IMPLEMENTATION

Fig 2. Shows the proposed architecture of CSNet. This architecture is basically divided into three parts, i.e. block-based compressed sampling, initial reconstruction, and non-linear signal reconstruction.

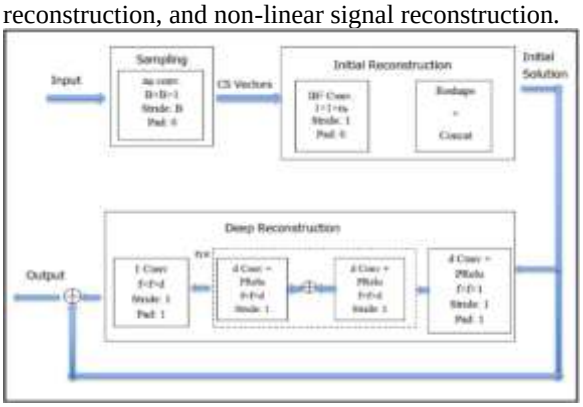


Figure 3 Flow Chart depicting the CSNet model architecture.

A. Sampling Network

In Block-based compressed sensing, the image is divided into $B \times B \times L$ where B is the height and width of the block and l denotes the number of channels. To make these blocks non-overlapping, we will be using a stride of Y.

The measurements are then obtained with a $n_b \times lB^2$ sampling matrix ϕ_B, where $n_B = \left[(M / N)^{lB^2}\right]$

$y_j = \phi_B X_j$ is the formula for this procedure.

If each row of the sampling matrix B is considered a filter, we may imitate this compressed sampling process using a convolutional layer.

Because each block of image is B × B × 1, each of the sampling network's filters is similarly B × B × 1, resulting in each filter producing one measurement.

To acquire n_B CS measurements, the sampling matrix B has n_B rows for the sampling ratio (M / N).

As a result, this layer has n_B filters of size B × B × 1. For non-overlapping sampling, this layer's Stride is B × B. Furthermore, each filter is free of prejudice. In addition, this layer has no activation[6].

The Block-based compressed sampling network (S) is conceptualized as

$$y = s(x) = ws * x$$

where * denotes convolutional operations, x denotes the input image, y represents the CS measurement, and $ws *$ denotes n_B filters that support B × B × 1. The result appears to be made up of n_B feature maps, with each column representing the n_B measurements of an image block.

B. Initial Reconstruction Network

The picture is rebuilt using a recovery network (abbreviated R), which consists of two networks. The first one is an Initial Reconstruction Network (abbreviated I) and the second one is a Deep Reconstruction Network (abbreviated D).

$$R(y) = D(I(y))$$

Provided the j^{th} block's CS measurement y_j, the first reconstruction yields $x_j\sim = \phi_B \sim yj$. Clearly, $\Phi^{\sim}_R$ is a matrix with dimensions of $1B^2 \times n_B$.

To implement the first reconstruction procedure, we employ a convolutional layer with a particular kernel size and stride. $\Phi^{\sim}_R$ is adaptively optimized in the network.

The first image reconstruction can be described as the following procedure I˜(y):

$$I\tilde{}(y) = W_{int} * y$$

The CS measurement is y, and the filters are W_{int}. The sampling network outputs a $1 \times 1 \times n_B$ vector for each image block and the size of the output of the sampling network and the size of each filter in the initial layer are the same.

As a result, W_{int} is equivalent to $1B^2$ filters with support of $1 \times 1 \times n_B$. To reconstruct each block, we have set the stride of this convolutional layer to 1 × 1. Bias isn't present either.

Each column of I˜(y) appears to be of 1 × 1× $1B^2$ type. The reassembled block, however, is still a vector. To get the initial linear reconstruction, the combinational layer has two functions, the first one is a reshape function and the second one is a concatenation function. This whole process has been denoted by *I(y):*

$x\tilde{} = I(y) = k$ *(for all image blocks* γ*(Image block))*

where γ is the reshape function and *k* is the concatenate function.[8][11]

The reconstruction of the signal in this network is linear as we are not using any activation functions

C. Deep Reconstruction Network

We have used residual block in this network to accomplish the non-linear signal reconstruction. There are three operations in this network: Feature extraction, non-linear mapping and feature aggregation.

D. Why have we used Residual blocks?

In a word, ResNets handle the problem of deep neural network effectiveness degrading, as the system grows larger.

A leftover block is a group of tiers in which the result within one layer is retrieved and combined with the result of a level farther in the block. The fluctuation is then transferred to the main path by mixing it with the result of the appropriate layer. This by-pass connection is known as a shortcut or skip-connection[4].

1) Feature Extraction:

The high-dimensional feature is extracted from the local receptive field using a feature extraction method. It consists of a convolutional layer and an activation layer.

The convolutional layer has d filters of size $f \times f \times 1$ since it works with the initial reconstruction output. This operation is denoted by the symbol $De(x\tilde{})$:

$$De(x\tilde{}) = Activation(W_e * x\tilde{} + B_e)$$

The initial reconstructed result generated by the Initial reconstruction network is denoted by $x\tilde{}$.

W_e stand for d filters with sizes of $f \times f \times 1$, B_e for biases of size $d \times 1$, and *Activation()* for activation function. For the activation function, we used the Parametric Rectified Linear Unit (PRelu, *max(0,x) + a * min(a,x)*).

2) Non-Linear Mapping:

This network dynamically cascades residual block convolutional layer and activation layer after obtaining the high-dimensional image feature, increasing the net-work non-linearity.

This procedure is written as:

$$D^i_{m1}(x\tilde{}) = Act\ (D^{i-1}_{m2}\ (x\tilde{}) + W^i_{m1} * D^{i-1}_{m2}\ (x\tilde{}) + B^i_{m1})$$
$$D^i_{m2}\ (x\tilde{}) = Act\ (\ W^i_{m2} * D^i_{m1}\ (x\tilde{}) + B^i_{m2})$$
$$D^0_{m2}\ (x\tilde{}) = D_e\ (x\tilde{})$$

Where i ∈ { 1,2,....,n }. There is a short skip connection between the input and output in the residual block. W^i_{m1} and W^i_{m2} contain $f \times f \times d$ size d filters, B^i_{m1} and B^i_{m2} are $d \times 1$ biases, and $Act(.)$ is a PRelu activation function.

3) Feature Aggregation:

A feature aggregation operation is used to reconstruct the image from the high-dimensional feature to provide the final output. The operation is described as $D_a(\tilde{x})$:

$$D_a(\tilde{x}) = W_a * D^n_{m2}(\tilde{x}) + B_a$$

W_a denotes a filter of size $f \times f \times d$, while B_a denotes the bias of size 1 X 1.

In the diagram, we can see that we have also made a long link between the linearly reconstructed picture and the deep reconstruction network's output $D_a(\tilde{x})$ to speed up network convergence. As a result, the final reconstructed image looks like this:

$$D(\tilde{x}) = \tilde{x} + Da(\tilde{x})$$

V. Result Analysis

Table 1 Results obtained from the deep learning model.

Number of Residual Blocks	PSNR
1	28.997789713705338
2	29.022606572723685
3	29.218752044341894
4	29.440226792141814
5	29.170336279307907

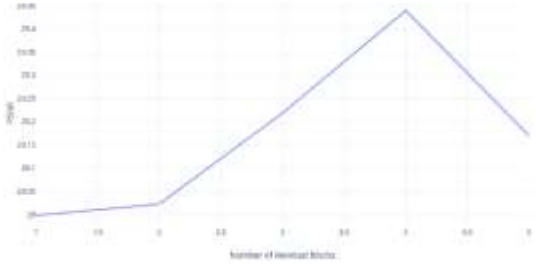


Figure 4 Plot of PSNR values vs number of residual blocks.

As shown in figure 4, initially the PSNR value rises with increase in the number of residual blocks, which implies that the quality of compressed images also increases. But after reaching a certain number of blocks (4 in this case), the PSNR value attains it's maximum value and decreases gradually thereafter. Therefore, 4 residual blocks is the most optimum setting to obtain high quality images from the CSNet framework.

VI. Conclusion & Scope

In this paper, a better approach to state-of-the-art image-compressed sensing techniques have been proposed. The traditional methods have many issues, like high computational power requirements and slow computational speeds.

Although there are methods that require less computation and produce faster results but at the expense of the quality of the output images. CSNet reconstructed images achieve high average PSNR values, and low computation times, depicting that the learned CS measurements retain more image structural information for better reconstruction, offering high quality compressed images at faster running speed. We found that the CSNet achieves peak results when four residual blocks were used.

Furthermore, there is a huge scope of research in the field of compressed data sensing. Like, applying the CSNet framework in real world applications such as facial recognition and MRI scans.

References

1. Wuzhen Shi, Feng Jiang, Shaohui Liu, Debin Zhao. "Image Compressed Sensing using Convolutional Neural Network", IEEE transactions on Image Processing, 2020.

2. P Hanumath, P Bhavana, Shreyanka Subbarayappa. " Application of deep learning and compressed sensing for reconstruction of images", Journal of Physics: Conference Series, 2020.

3. Yi Xu, Guiling Sun, Tianyu Geng, Ying Zhang. "A two-Stage Fusion Method of Reconstruction Algorithms for Compressed Sensing", International Journal of Wireless Information Networks, 2018.

4. Su H Bao,Q,Chen Z, ADMM Net " A Deep Learning Apporach for parameter Estimation of Chirp Signals under Sub Nyquist Sampling", IEEE Acess 2020,8,75714-75727.

5. Zheng, S.; Yang, X.; Yuan, X. Two-Stage is Enough: A Concise Deep Unfolding Reconstruction Network for Flexible Video Compressive Sensing. arXiv 2022, arXiv:2201.05810.

6 K Kulkarni, S Lohit,P Turaga, R Kerviche, A.Ashok " Recon Net Non iterative reconstruction of images from compressively sensed measurements", Proc IEEE Conf Vis Pattern Recongit PP 449-458 Jun 2016.

7. W. Shi, F. Jiang, S. Zhang and D. Zhao, "Deep networks for compressed image sensing", *Proc. IEEE Int. Conf. Multimedia Expo (ICME)*, pp. 877-882, Jul. 2017.

8 Y. Wang, W. Yin and J. Zeng, "Global convergence of admm in nonconvex nonsmooth optimization", *Comput. J. Sci. comput.*, pp. 1-35, 2018.

9 K. Zhang, W. Zuo, L. Zhang et al., "Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising", *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142-3155, Jul. 2017.

10 D. L. Donoho, "Compressed sensing", *IEEE Trans. Inf. Theory*, vol. 52, no. 4, pp. 1289-1306, Apr. 2006.

11 L. Zhu, X. Wu, Z. Sun, Z. Jin, E. Weiland et al., "Compressed-sensing accelerated 3-dimensional magnetic resonance cholangiopancreatography: Application in suspected pancreatic diseases", *Invest. Radiol.*, vol. 53, no. 3, pp. 150-157, 2018.

12 S. Beygi, S. Jalali, A. Maleki and U. Mitra, "An efficient algorithm for compression-based compressed sensing", 2017.

13. S. Ravishankar, B. E. Moore, R. R. Nadakuditi and J. A. Fessler, "Low-rank and adaptive sparse signal (LASSI) models for highly accelerated dynamic imaging", *IEEE Trans. Med. Imag.*, vol. 36, no. 5, pp. 1116-1128, May 2017.

The International Journal of Advanced Manufacturing Technology
https://doi.org/10.1007/s00170-022-09324-8

ORIGINAL ARTICLE

Implementation of mass customization for competitive advantage in Indian industries: an empirical investigation

Piu Jain[1] · **Suresh Garg**[1] · **Gayatri Kansal**[2]

Received: 3 August 2021 / Accepted: 2 May 2022
© The Author(s), under exclusive licence to Springer-Verlag London Ltd., part of Springer Nature 2022

Abstract
The enduring fluctuations in market demand, exemplified by exceedingly unpredictable customer requirements, have given rise to mass customization (MC), which is acquiring increasing prominence in production and operations management. Fostering on the foundation laid by erstwhile researcher Hart in Int J Serv Ind Manag 6:36–45 [1], who developed an analytical framework of four pillars of mass customization for organizations, the objective of this research is to obtain additional discernments on the nature of the linkage between the four pillars and MC. The current work is to explain ability of manufacturing organization for MC. Further, the paper proposes the impact of MC on organization performance with a view to develop an assessment and decision-making model to achieve competitiveness. The analysis of 276 valid responses from Indian professional experts was done using structural equation modeling (Amos Graphics). The results showed that process amenability, competitive environment, and organizational readiness all have a major impact on mass customization. The study, on the other hand, did not support customer sensitivity, which could be because mass customization is a new manufacturing paradigm in India.

Keywords Mass customization · Competitive advantage · Manufacturing · Structural equation modeling · Indian Industry

1 Introduction

Economic turmoil, global value chain disruptions, impulses of nationalism, and protectionism have all been exacerbated by the onset of COVID-19, which has heightened the challenges to globalization and the international business arena [2]. To ensure competitiveness, the advancement of a post-COVID-19 business model would necessitate confronting challenges and will necessitate rethinking [3]. Given the circumstances, MC requires adopting methods and developing a manufacturing framework that supports flexibility, quick customer response, cost-competitive products, product design tailored to customers' needs, and achievement of corporate objectives [4].

Several researchers have characterized mass customization as a production paradigm endowing profound strategic opportunities [5] with the potential of managerial response during a progressively uncertain, competitive, and intricate environment [6], having a competitive positioning strategy in the marketplace [7] during the period exemplified by the introduction of innovative production technologies, increased global competitiveness, reduced product life cycles, and the customer need for more product diversity [8]. Organizations aiming to mass customize, on the other hand, will have hurdles in designing a system capable of acquiring and analyzing a wide range of complex and uncertain data [9]. The ability to manufacture mass customized goods is dependent on the ability of a corporation to use a range of production issues simultaneously, including quality, inventory, process technology, and human skills [5]. For effective application of their MC ability for competitive advantage, organizations must identify and exploit resources and skills effectively and efficiently [10]. This necessitates the development of a manufacturing model to identify the factors a manufacturer needs to consider for the effective implementation of MC in an organization.

✉ Piu Jain
piujaink@gmail.com

Suresh Garg
skgarg63@yahoo.co.in

Gayatri Kansal
gayatri@kenstel.com

[1] Delhi Technological University, New Delhi, India

[2] Department of Mechanical Engineering, School of Engineering and Technology, IGNOU, New Delhi, India

Published online: 20 May 2022

Most of the research to enhance MC ability and improve an organization's MC capabilities was carried out in the areas of organizational design and manufacturing practices or product and customer aspects like quality management [11], product modularity [12], functional integration [13], information technology [14], work-design practices [15], operations and marketing functions [16], supply chain integration [17], organizational learning and process technology [18], organizational structure [19], CAD/CAE integrated customization product approach [20], product modularity on supply chain quality integration [21] customer integration [22], and degree of consumers involved in the value chain [23]. Some studies identified technology needed like Industry 4.0 [24], RFID-enabled real-time manufacturing execution system [25], additive manufacturing [26], and computer-aided manufacturing planning [27]. None of these preceding research analyzed the role of operational agility required for MC ability. Organizations entailing to implement mass customization to enhance market competitiveness require to improve the operational agility of the MC production model to increase the flexibility, speed, and efficacy of responding to unique client needs [28].

Some research identified the role of enablers in manufacturing [29, 30], but used ISM technology which is solely dependent on the judgment and qualitative inputs of a few experts and was used to build theory, and test theory using structured questionnaires, and in a single analysis, estimates of multiple and linked dependency were needed. Research work conducted was region-specific or product-/sector-specific, or both like Chinese automotive suppliers [31], door sector in Poland [32], pigment company in Taiwan [33], and shoe manufacturing in India [30].

Research revealed that mass customization is an incipient concept in the Indian industry [30]; hence, organizations of Indian origin are apprehensive of their technological potential and are not taking initiative to transform themselves from conventional mass manufacturing companies to mass customization for emerging markets' needs [34]. There is very little effort done by such industries on implementing mass customization as a manufacturing strategy and developing a roadmap for MC adoption. To comprehend and facilitate their transition to mass customization, it is required to build strategic roadmaps that specify their future directions [35], since it is agreed by many researchers that mass customization is a feasible way to competitive strategy. Based on the above research study, the following gaps were observed:

RQ1: *What factors affect the adaptation of mass customization during the transition from mass production to mass customization?*

RQ2*: What is the impact of these factors on competitive advantage concerning the manufacturing sector of Indian origin?*

Manufacturing sectors in India identifying the possibility of moving away from mass manufacturing and toward mass customization require a manufacturing model to enhance competitive advantage and tide over competitors and needed to identify operational agility to enhance MC ability. Based on the above research questions, the following objectives have been defined for the present study:

- To undertake an extensive literature review to develop a framework that will facilitate mass customization implementation.
- To identify factors that influence enterprises in the adaptation and enhancement of mass customization ability, specifically in the Indian economic condition and to statistically validate the framework.
- To identify the theoretical and managerial implications of the research work.

The goal of this research is to empirically resolve gaps in MC theory, with an emphasis on how customer, process, competition, and organizational factors affect mass customization capabilities for competitive advantage in India's manufacturing industry. The four pillars of mass customization established by Hart (1995) were further investigated using structural equation modeling (SEM), which was utilized to evaluate the data from 276 manufacturing enterprises of Indian origin.

The paper is organized as follows: Sect. 2 represents the theoretical background and hypothesis development. Section 3 highlights the research methodology. Section 4 presents analysis and results. Section 5 articulates the results and discussion. The managerial implication of the study is provided in Sect. 6. Section 7 presents the conclusion. Section 8 conveys the limitation of the study and future research direction.

2 Theoretical background and hypothesis development

The research is based on developing a manufacturing model for mass customization ability, as all operations in the product design, manufacturing, and delivery processes must be better coordinated and integrated [13]. To solve internal and external supply chain quality challenges, managers must use manufacturing and organizational design principles to improve mass customization capabilities [21]. Suppliers and customers' amalgamation

has a major impact on the agility and flexibility of an organization, which is necessary for MC capabilities [36]. Tookanlou and Wong [37] identified mass customization as a promising technique that has grabbed the interest of practitioners and researchers due to its potential to help firms gain a competitive advantage, generate profits, and minimize waste through on-demand production. The hypotheses offered in this study are based on the four pillars of mass customization proposed by Hart (1995) and further strengthened by investigating the relationship between MC ability for competitive advantage from the extant literature. The constructs, related to measurement items, are presented in Table 1.

2.1 Customer sensitivity

The uniqueness of customers' demands and consumer sacrifice for unmet needs are the primary drivers of customer customization sensitivity [1]. To create mass customization capability, organizations must improve their operational

Table 1 Constructs, measurement items with factor loads

Constructs	Measurement items	Factor loads	Cronbach's alpha (> 0.70)	Item R-square	CITC (> 0.30)
Customer sensitivity	**CS1** Identify opportunities for customization	0.902	0.958	0.814	0.714–0.874
	CS2 Understand the uniqueness of customers' needs and create value	0.927		0.859	
	CS3 Analyze customers sacrifice for unmet needs	0.845		0.714	
	CS4 Provide ease of customer choice for decision making	0.883		0.780	
	CS5 Create value for the customer	0.926		0.857	
	CS6 Incorporate customer requirement during new design process	0.865		0.748	
Process amenability	**PA1** Incorporate modularity in design, for part flexibility	0.825	0.924	0.681	0.635–0.784
	PA2 Develop compatible process technology	0.765		0.585	
	PA3 Develop production and distribution process for timely delivery	0.881		0.776	
	PA4 Develop supplier for co-design	0.872		0.760	
	PA5 Develop marketing competence for product promotion	0.863		0.745	
Competitive environment	**CE1** Economic uncertainty	0.771	0.917	0.594	0562–O.835
	CE2 Market Turbulence	0.758		0.575	
	CE3 Company credibility and position in marketplace	0.851		0.724	
	CE4 Customer loyalty	0.868		0.753	
	CE5 Potential of competitors to react	0.873		0.762	
Organizational readiness	**OR1** Cultural change in the organization	0.866	0.954	0.750	0.754–0.867
	OR2 Skill development of employees	0.924		0.854	
	OR3 Employee's involvement in product and process roadmap	0.869		0.755	
	OR4 Training to marketing team to capture and prioritize requirements	0.914		0.835	
	OR5 Top management support and leadership building	0.918		0.843	
Mass customization ability	**MC1** Product can be customized on large scale	0.848	0.908	0.719	0.647–0.782
	MC2 Product variety can be enhanced at same cost and quality	0.833		0.654	
	MC3 Product can be designed based on customers' requirements	0.862		0.740	
	MC4 Product can be delivered on stipulated time	0.830		0.689	
Competitive advantage	**CA1** Market share growth and Reaching financial goals	0.815	0.906	0.664	0.584–0.703
	CA2 Acquiring new customer	0.812		0.659	
	CA3 Increased sales volume/ Return on sales/ revenue	0.846		0.716	
	CA4 Increased product variety	0.783		0.613	
	CA5 Reduce waste through on demand production	0.813		0.661	

efficiency, which necessitates the interchange of information between customers, salespeople, and technical personnel [38]. Customers must transform their wants and demands into particular product requirements [22] by customizing, setting, matching, or altering a solution, limiting the solution space to options already represented in the system's fulfillment, and converting consumer co-design data into customer knowledge for strategic planning and innovation [39, 40]. Customer input to establish features and pricing requirements of personalized products to meet consumers' particular requirements cheaply and promptly [41], and a study of consumer buying patterns [42] can contribute to success in this field. Quick customer response for the implementation of mass customization is built on a system that connects consumer voice to product design, customer relationship management, regular market surveys, and understanding client grievances for product improvement [34]. Thus, the following hypotheses were proposed:

Hypothesis H1: *Customer sensitivity is positively related to MC ability.*

2.2 Process amenability

Enablers, marketing and strategy, design, production, and distribution are all part of this diverse region [1]. New technology must be integrated and coordinated with humans and organizations to successfully execute mass customization [43]. Flexible manufacturing techniques are required by MC to help businesses become first to market with unique, custom-made items and [44] to manage the complexity of mass customization [45]. Increasing product variety implies richer information flows, which in turn necessitates increased scheduling and coordination complications inside the organization [13, 46], giving rise to increased information processing needs [13]. Organizations need to establish suitable information infrastructure to expedite regular communication and information with suppliers, customers, and other stakeholders [29, 43, 47]; to keep close communication with its supplier for design consideration [48]; and to determine interface specifications to ensure module combinability [14]. Encompassing suppliers in quality enhancement and new product improvement helps a company to tap into their capabilities and competencies, which can moderate new product development expenses and lead times while also minimizing component mismatches [49, 50]. Collaboration with suppliers for new product development and standards by firms is vital since modules require consistent specification [21]. Modular product design is based on modular components that may be assembled into several end products [51] as well as the reuse of the same modules across several end products [52], resulting in increased product volume and variety [12]. Modularity primarily influences the desirability of products by developing tenuously connected modules that may be procured from vendors [32] and merged according to consumer requirements and manufacturing processes [53]. It is critical to maintain constant, direct contact with customers and involve them in the design process to incorporate values that are important to them into the design [32]. Managers, according to Zhang et al. [41], must communicate with consumers and suppliers at the same time to incorporate MC concepts throughout the design phase. The following hypotheses are based on the previous reasoning.

Hypothesis H2: *Process amenability is positively related to MC ability.*

2.3 Competitive environment

Competitive environment, economic uncertainty, market turbulence, firm credibility, and position in the marketplace, as well as consumer loyalty, are all elements to consider in the journey to mass customization [1]. Researchers have advocated that enterprises that change from mass production to the novel paradigm of mass customization will acquire a competitive benefit [8] due to the growing awareness of mass customization. Organizations require a greater diversity of supply to meet heterogeneous and changeable consumer demand, which upsurges the ambiguity in forecasting demand for each type of product. This causes scarcities, quality concerns, and extended lead times in supplier parts, which causes manufacturing delays and limits a company's capacity to mass customize [13]. To successfully adapt to altering market wants, MC capability development necessitates extraordinary levels of process flexibility and agility inside a company [54], as well as excellent internal integration across multiple roles. The conception of mass customization has been in practice to improve an organization's ability to respond fast to dynamic changes in the global marketplace [27]. This method cannot only improve resilience, preserve operations, and support personnel during economic downturns, but also maintain a competitive advantage and accelerate business growth [34]. The following hypothesis is proposed based on the above research findings:

Hypothesis H3: *Competitive environment is positively related to MC ability.*

2.4 Organizational readiness

Organizational readiness demands a thorough examination of the organization's approaches, culture, and resources [55] to determine the extent of similarity between the commercial opportunity offered by mass customization and the organization's capacity to profit from it [1]. To attain new and innovative forms of competitive advantage, enterprises must integrate, build, and reconfigure available

resources to satisfy the expectations of a continually changing environment [17]. Before transitioning from a conventional mass production setup to a successful MC firm, enterprises must make significant changes to their historical mindsets and practices [56], since mass customization involves true change management initiatives in existing enterprises [39]. To resolve the hindrances faced by organizations while implementing mass customization, intensification of the training of designer and personnel reassignment needs to be considered for the implementation of MC ability and enhancing their market competitiveness [28]. Organizational performance necessitates a group of workers with cross-functional skills to foster creativity and innovation throughout the entire value chain to improve MC's ability to achieve organizational goals [29]. Mass customization is presented as a marketing concept that allows for the acquisition of a large number of customers while providing the opportunity to personalize the product [57]. Hence, organizations require to bring a cultural change, involve their workforce in all phases of MC implementation, upgrade and train employees for MC innovative ideas, train their marketing team to understand the desire of customers, and market requirements that need to be converted into various MC aspects of design. Such organizational readiness brings about operational agility for MC adaptation and implementation. Organizations having unique operational capabilities with collaborative relationships with the suppliers, flexible, innovative, and adaptive work culture have strong MC ability [29]. Tung et al. [58] emphasized that top management's participation in problem resolution was crucial during crisis and conflict; therefore, all across the execution phase, it was essential to simultaneously enhance production, process, and product on the strategic and tactical level, which might be a challenging assignment for top management [42]. Effective leadership for analyzing future market potential, participation to ensure technical, economic, and personnel support, aid during obstructions, and strategic monitoring of MC projects and appraisal of their progress are the most important criteria for mass customization success [34]. Organizations that are flexible ensure mobility, agility, and adaptability and are more capable to reduce the response time to demand changes [59]. Companies that are self-sufficient, self-organized, linked with intelligent digitalization and forms of communication, and seem to have independent entity management teams utilize additional supplier networks, guarantee better supply chains' capacity and access to external resources, have a protocol in place and diverse supply alternatives, and facilitate collaboration among supply chains that are more ready for MC implementation. Thus, the following hypotheses are proposed:

Hypothesis H4: *Organizational readiness is positively related to MC ability.*

2.5 Mass customization ability

MC is becoming a more important strategic goal as competition grows and customers become more assertive [19]. The ability to manage change has the potential to improve operational, market, environmental, and financial performance [57]. When faced with fierce competition, many businesses want to improve their MC skills to outperform their rivals [6]. An organization's effective application of mass customization gives a competitive edge in the industry [54] and aligns manufacturing with customer needs, satiating client desires [60, 61], which is regarded as a key aspect of any organization's competitiveness [21, 57, 62]. Manufacturing agility can improve organizational performance by generating operational efficiency and strategic competitive advantages [60]. If each product's originality or clarity of innovation is maintained, a diverse range of options stimulates customers and opens the door to increased sales, profit margins, and revenues [44, 63]. The following hypothesis is proposed based on the previous arguments:

Hypothesis H5: *Mass customization ability is positively related to competitive advantage*

2.6 Research model

Figure 1 illustrates the proposed research model, the hypothesized relationship, and the equivalent which has been previously explained in the hypotheses development section.

3 Research methodology

The main goals of this study are to see how different constructs stated by Hart (1995) affect the ability of an Indian company to mass customize products, which is necessary for competitiveness.

3.1 Design of survey instrument and its reliability

The four pillars of mass customization developed by Hart in 1995 were adopted as the manufacturing model based on the research gap, and a preliminary collection of several factors was generated based on a literature review, experience surveys, in-depth interviews, focus groups, and critical incidents [64]. The next step was to create a questionnaire based on a literature study, with each latent variable analyzed consisting of a set of items to be evaluated. There were three sections to the questionnaire. The first component of the questionnaire covers questions about the respondents' demographics, the sort of industry

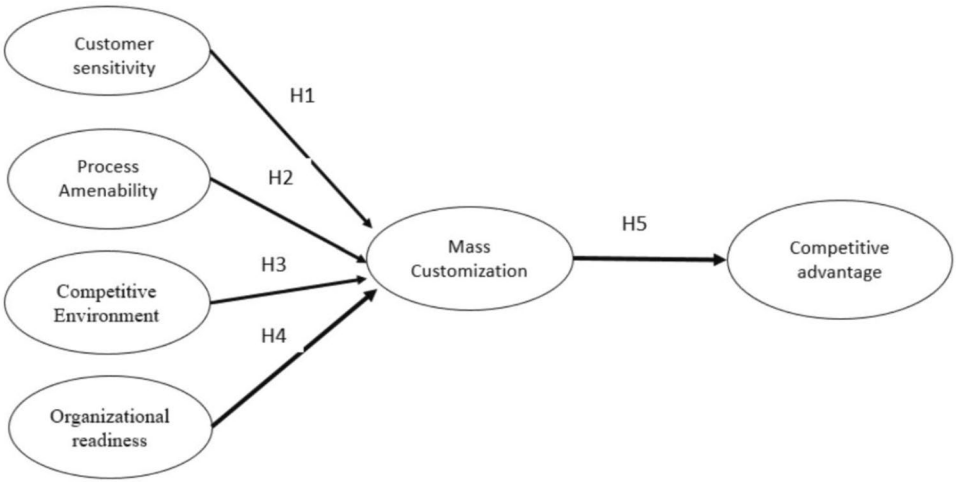


Fig. 1 Research framework

in which they work, and the size of their company in terms of manpower and turnover. The next section contains questions about the respondents' perceptions of their level of MC skill. The final segment contains questions about how respondents feel about various aspects of MC. The response format was a five-point Likert scale, with values ranging from 1 to 5, with 1 indicating strong disagreement and 5 indicating strong agreement [65, 66].

The questionnaire was subjected to a panel examination to ensure that it was accurate and clear [67]. This panel consisted of ten experts who were specifically briefed about the research concept and its dimensions and was finalized by engaging individually with six experts from manufacturing firms who were managers, vice presidents, general managers, assistant general managers, and other positions; two consultants from the field with extensive experience; and two academic experts (Ph.D. holders from prestigious universities). The expert screening process was developed in such a way that the panel of evaluators could discover items that were too identically worded [68]. The updated pool of topics was validated with a smaller group of respondents after the questionnaire was validated to ensure that all of the items were clear and understandable to the target audience [69]. The next step was scale purification, which involved using a refined and reduced scale for data collection while adhering to a sampling strategy that was adequately justified in the context of the study.

3.2 Target organizations and target respondents

Respondent lists were obtained from reputable sources such as chambers of commerce and industry, trade organizations, and alumni associations. The most important criteria, limits, and obstacles were to identify respondents with prospective expertise to address knowledge in the questionnaire from a variety of fields covered by various domains [70]. The survey was conducted in India's manufacturing and processing industries. Companies that used or were familiar with such technology were chosen as respondents.

3.3 Collection of data

To obtain responses from potential respondents, the questionnaire was distributed both offline and online. In the case of offline mode, respondents were requested to schedule an appointment ahead of time. A soft copy of the questionnaire was then emailed to them, along with a supplemental background note on mass customization, to help them better comprehend the many aspects. A face-to-face interview was used to obtain responses from the respondents. In the instance of an online survey, participants were asked to make their beneficial inputs available. The questionnaire and background note on MC were emailed to 856 respondents from India's manufacturing businesses, representing a variety of industries. The filled-in surveys were examined to see if the respondents gave significant inputs once the online and offline responses were received.

3.4 Respondent characteristic

The questionnaire and background note on MC were emailed to 856 respondents from India's manufacturing businesses, representing a wide range of industries. The filled-in questionnaires were examined after the online and offline responses were collected to assess if the respondents made relevant contributions. Out of the 856 surveys distributed, 286 were completed and returned. In the surveys, missing data and outliners were examined [71]. For 10 surveys, outliners were recorded, generating 276 acceptable responses. The final response rate was 32.24%, which was higher than the minimum of 20% proposed by Malhotra and Grover [72] for guaranteeing the quality of the empirical research.

Sample characteristic of the respondents exhibits that as per job designation, top management constituted 34.6%, middle management 42.9%, administrative staff 8.3%, technical staff 3.9%, academician 10.2%,marketing professionals accounted for 17.3% of the respondents, followed by production 16.9%, procurement 15.4%, logistics 11.4%, R&D 10.2%, and project management 8.3%. In terms of time spent in the company, 5.1% of those polled have worked for their businesses for fewer than 5 years. 4.3% of respondents have worked for the company for 5 to 10 years and 23.2% for 10 to 20 years, and 49.6% have worked for the company for more than 20 years. Respondents had a 3.9% post-doctoral, 18.5% Ph.D., 55.1% postgraduate, and 17.9% graduate, and the rest are diploma holders as their educational backgrounds.

The sample characteristic of the surveyed organization consists of consumer electronics 18.5%, interiors and decors 18.1%, automotive 14.2%, apparel and footwear 9.8%, and food 9.1%, and the rest were variable industries. Concerning the strength of workers, organizations with more than 1000 workers were 32.7%, 1000 to 500 was 25.6%, 200–500 was 18.5%, 100–200 was 11%, and less than 100 was 12.5%. Annual turnover in crores for less than 10 was 21.7%, 10 to 50 was 15.4%, 50 to 100 was 8.5%, 100 to 500 was 17.3%, and more than 500 was 35.8.

3.5 Tests for potential bias in survey data

The responses were examined for non-response bias, which might reduce response validity (Bailey, 1978). The application of *t* tests to analyze non-response bias is a common strategy, assuming that late responders (e.g., second-wave replies) can be seen as non-responses of early responders (e.g., the first wave). For this study, 20 survey items were chosen at random for analysis, two groups of 50 surveys were picked at random from the first and last waves of surveys received, and the results of the two groups were compared using *t* tests [73] with the help of SPSS program. Using *t* tests, there were no statistically significant differences between the 20 survey questions. Although these findings do not rule out the possibility of non-response bias, they do show that non-response may not be a problem if late responders represent the views of non-respondents. As a result, scale purification and a confirmatory factor were added to the data analysis.

Because the data came from a single survey, there was a risk of common method bias [74], so Harman's single-factor test was run on the variables using SPSS and exploratory factor analysis (EFA). The findings revealed that a single component was responsible for 48.25% of the total variance, which is less than 50%. The findings indicate that no one or general factor appeared, and it was considered that common technique bias was not a concern.

4 Analysis and results

It is critical to ensure that the given theoretical framework is reliable and valid before testing it ([73, 75]).

4.1 Reliability and validity analysis

Both EFA and CFA are used to test the concept's reliability and validity. To examine construct reliability, EFA is used to test the scales' unidimensionality, followed by Cronbach's alpha and composite reliability (CR) [75]. The correlated item–total-correlation (CITC) was calculated to do the reliability analysis. A CITC value of more than 0.30 is recommended. Table 1 illustrates the CITC. As indicated in Table 1, all six constructions had CITC values greater than 0.30 [76]. As a result, the scales are determined to be internally reliable, and the construct dependability of the measurement model is not impacted.

EFA with principal component analysis and varimax rotation with Kaiser normalization is carried out [41] using SPSS. The Kaiser Meyer-Olkin value of 0.941 is higher than the suggested minimum value of 0.5 [77], indicating that the sample size is sufficient. The Bartlett test of sphericity, which is a measure of the multivariate normality of a set of distributions and evaluated the null hypothesis in SPSS, was used to determine the strength of the link [77]. Table 2 shows how six EFA components affect the loadings of measuring scale variables. It meets the requirement of clarifying the factors and the hypothesized model based on theory because components 1, 2, 4, and 5 correspond to the four pillars of mass customization, namely, customer sensitivity, organizational readiness, competitive environment, and process amenability, and components 3 and 6 correspond to competitive advantage and mass customization ability. The variable loading ranges from 0.689 to 0.886. For each construct, a Cronbach's alpha value was calculated, which ranged from 0.90 to 0.95 (Table 1), all of which are higher than the recommended threshold value of 0.70 [76].

Figure 2 shows a CFA model with the AMOS 23 program. Each item in the model is associated with a construct, and the covariance between those constructs is computed. The data were examined for skewness, kurtosis, and normality. There was no major violation found. According to the modification indices, no major changes were required for this model. The chi-square statistic to degrees of freedom (CMIN/DF) ratio, adjusted goodness of fit (AGFI), goodness-of-fit index (GFI), root mean

Table 2 Rotated Component Matrix[a]

	Component					
	1	2	3	4	5	6
Create value for customer	.880	.156	.140	.113	.218	.084
Understand uniqueness of customers' needs	.840	.161	.171	.211	.200	.167
Provide ease of customer choice for decision making	.832	.162	.196	.150	.201	.082
Identify opportunities for customization	.832	.128	.160	.208	.216	.110
Incorporate customer requirement during new design process	.818	.092	.150	.164	.268	.043
Analyze customers sacrifice for unmet needs	.794	.185	.167	.220	.164	.087
Skill development of employees	.156	.865	.206	.140	.164	.132
Employees involvement in product and process roadmap	.134	.855	.173	.075	.135	.128
Training to marketing team to capture and prioritize requirements	.213	.851	.200	.134	.129	.123
Cultural change in the organization	.100	.848	.185	.178	.109	.110
Top management support and leadership building	.181	.843	.213	.129	.174	.158
Increased sales volume/ Return on sales/revenue	.189	.164	.822	.119	.094	.152
Market share growth/Reaching financial goals	.189	.264	.765	.135	.083	.128
Reduce waste through on time production	.139	.183	.758	.154	.188	.217
Acquiring new customer/Perceived customers value	.180	.138	.753	.177	.217	.190
Increased product variety	.159	.242	.753	.127	.016	.186
Market Turbulence	.178	.084	.129	.834	.202	.169
Economic uncertainty	.211	.065	.135	.792	.195	.166
Potential of competitors to react	.190	.250	.160	.742	.241	.197
Customer loyalty	.301	.214	.178	.679	.203	.218
Company credibility and position in marketplace	.217	.196	.246	.677	.313	.205
Develop marketing competence for product promotion	.250	.152	.168	.279	.765	.158
Develop production and distribution process for timely delivery	.317	.208	.121	.203	.746	.235
Develop supplier for co-design	.268	.173	.145	.295	.735	.206
Incorporate modularity in design, for part flexibility	.332	.149	.060	.267	.715	.217
Develop compatible process technology	.286	.176	.196	.182	.686	.188
Product can be customized on large scale	.096	.183	.174	.245	.165	.817
Product variety can be enhanced at same cost and quality	.078	.168	.192	.236	.224	.786
Product can be designed based on customers' requirements	.161	.159	.266	.211	.227	.761
Product can be delivered on stipulated time	.158	.157	.398	.154	.249	.684

Extraction Method: Principal Component Analysis

Rotation Method: Varimax with Kaiser Normalization

[a] Rotation converged in 6 iterations

square residual (RMR), normed fit index NFI, Bentler comparative fit index (CFI), and root mean square error are all indicators of acceptable fit [78] and the adequate fit suggested for models is shown in Table 2. It can be seen from Table 3 that all the measurement models have acceptable fit indices, and consequently signify the unidimensionality of the constructs.

After the factors are extracted using EFA and the quality of fit is confirmed using CFA, validation is done in two steps. Expert opinion is used to assess content validity, and construct validity is divided into two parts: convergent validity and discriminant validity. The values of CR (construct reliability) and AVE (average value efficiency) determine convergent validity (average value). The CR and AVE values should both be more than 0.7 and 0.5, with CR always being bigger than AVE [79]. Discriminant validity is determined using MSV values (mean shared variance) [78]. The AVE value should always exceed the MSV value. Table 4 backs up the

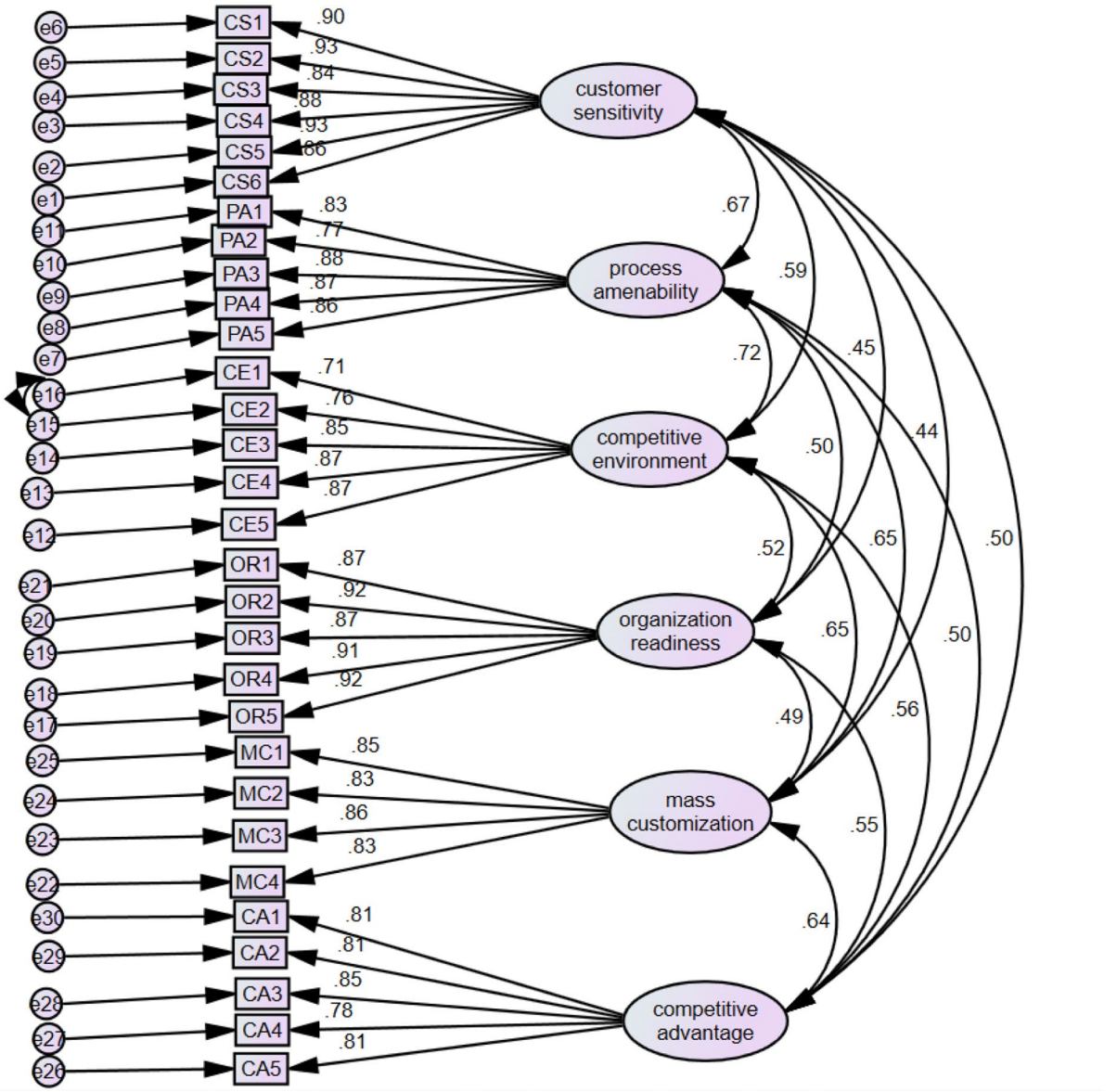


Fig. 2 Path diagram showing the regression weights and the correlation between the constructs

findings, and both convergent and discriminant validities exist.

4.2 Means, standard deviations, and correlations

To acquire a thorough grasp of the data, descriptive statistical analysis was used. As a result, descriptive statistics and bivariate correlation data derived from the Pearson coefficient correlation approaches have been assessed (Table 5). Table 5 reveals that the correlation coefficients for the constructs are greater than 0.40 at the 0.01 level (2-tailed), showing that they are positively and statistically significantly correlated [80]. However, because no relationship is greater than 0.70, multi-collinearity is unnecessary [80].

Table 3 Model fit indicators for CFA

Indicators of model fit	CMIN/DF	AGFI	GFI	RMR	NFI	CFI	RMSEA
Result	1.73	0.83	0.86	0.032	0.91	0.96	0.050
Default value	Less than 2	Greater than 0.8	Greater than 0.8	Less than 0.05	Greater than 0.9	Greater than 0.9	Less than 0.06

Table 4 Assessment of Discriminant Validity and Convergent Validity

	CR	AVE	MSV	ASV	Customer Sensitivity	Process Amenability	Competitive Environment	Organizational Readiness	Mass Customization	Competitive Advantage
Customer Sensitivity	0.96	0.80	0.45	0.28	**0.89**					
Process Amenability	0.92	0.71	0.52	0.37	0.67	**0.84**				
Competitive Environment	0.91	0.66	0.52	0.37	0.59	0.72	**0.81**			
Organizational Readiness	0.95	0.81	0.27	0.25	0.45	0.50	0.52	**0.90**		
Mass_Customization	0.90	0.71	0.42	0.33	0.44	0.65	0.65	0.49	**0.84**	
Competitive_Advantage	0.90	0.66	0.41	0.30	0.45	0.5	0.56	0.55	0.64	**0.81**

4.3 Hypothesis testing

Using AMOS 23, SEM analysis is used to analyze the associations between each pair of postulated components, as shown in Table 6. Figure 3 shows a path diagram created with the AMOS software., which portrays the results of the analysis derived from the SEM. Model fit indices values show good fit (CMIN/DF = 1.83; GFI = 0.85; AGFI = 0.83, NFI = 0.91, CFI = 0.95, RMSEA = 0.05). All these values are satisfactory for the recommended range values.

In terms of the impact of customer sensitivity on MC ability, the SEM results do not support H1, and the path coefficient is $\beta = -0.05$, $t = -0.096$, thus disagreeing with the suggestions of Hart (1995) that the customers' needs and wants and his sacrifice for unmet needs require to be considered by any organization before shifting from mass production to mass customization. The result is surprising, but it may be because the analysis was carried out in India where mass customization is at its nascent stage, and respondents' ability to analyze the gap between products provided and products desired by the customer gives rise to mass customization ability was crucial. As hypothesized, a significant relationship between process amenability and mass customization ability is observed ($\beta = 0.28$, $t = 4.07$, significant at $p < 0.00$), thus supporting H2. This concurs with the research by Hart (1995) that organizational enablers, marketing competence, supportive production and distribution team, and design aspects can enhance the mass customization ability of a firm. Hypothesis H3 is supported ($\beta = 0.34$, $t = 4.32$, significant at $p < 0.00$), acceding Hart's research theory that a competitive environment promotes mass customization. Hypothesis H4 showing organizational readiness leads to mass customization ability is supported ($\beta = 0.14$, $t = 3.12$, significant at $p < 0.002$), harmonizing with research by Hart (1995) that organizations need to identify the fit between business opportunity and organizations' ability to capitalize on this opportunity before venturing into mass customization. Finally, H5 is supported ($\beta = 0.69$, $t = 10.23$, significant at $p < 0.000$), emphasizing that MC-capable manufacturing plants may dynamically modify their resource/skill mix to adapt to individual client demands, gaining a competitive advantage [13].

5 Results and discussion

To meet customer demand in a certain period, a manufacturing company must choose the best acceptable product configuration from a wide range of options to achieve effective MC, which in turn relies on the firm's capacity to effectively grasp client wants and obtain the necessary mix and quantity of components for timely assembly of the needed product configuration. The goal of this study is to categorize, consolidate, and validate important factors that influence the ability of Indian manufacturing companies to integrate mass customization. Hart's (1995) four pillars of mass customization were used to identify constructs, which are then followed by a systematic literature review to identify measurement items. The study

Table 5 Descriptive analysis and correlations

	Mean	SD	Customer Sensitivity	Process Amenability	Competitive Environment	Organizational Readiness	Mass Customization	Competitive Advantage
Customer Sensitivity	3.74	0.96	**1**					
Process Amenability	3.79	0.88	0.67	**1**				
Competitive Environment	3.80	0.79	0.59	0.70	**1**			
Organizational Readiness	3.87	0.89	0.45	0.50	0.52	**1**		
Mass_Customization	3.95	0.75	0.44	0.65	0.65	0.49	**1**	
Competitive_Advantage	3.88	0.72	0.45	0.50	0.56	0.55	0.64	**1**

Table 6 Result of hypothesis testing

	Hypothesis	*B*	C.R (β)	*p* value	Supported
H1	Customer sensitivity is positively related to MC ability	−0.05	−0.096	0.33	Not supported
H2	Process amenability is positively related to MC ability	0.28	4.07	0.00	Supported
H3	Competitive environment is positively related to MC ability	0.34	4.32	0.00	Supported
H4	Organizational readiness is positively related to MC ability	0.14	3.12	0.002	Supported
H5	Mass customization ability is positively related to competitive advantage	0.69	10.23	0.00	Supported

used this information to describe the links between mass customization and competitive advantage, as well as give empirical evidence for the proposed research technique.

Due to its cost competitiveness, qualified labor, and favorable government regulations, India has emerged as a worldwide manufacturing hub. Furthermore, the presence of a robust local market is the most important factor promoting sector growth. The country's consumer trend is allowing domestic players to thrive, while also enticing international players, and the Indian market has a plethora of unexplored potential. Though literature support has been observed globally in the electronic sector [81], in the automotive industry [82], the fashion industry [83, 84], footwear [85], pigment industry [33], door industry [32], and food industry [56], in the Indian business, MC is still an emerging paradigm. Research work identified MC adaptability in the footwear industry [30], pigment industry, furniture, and electrical sector [34] in the Indian context, thus identifying with the view that though there is a lot of literature detailing the approaches and capabilities needed for MC system implementation, research in Indian manufacturing sectors is needed to identify the challenges like new system design,

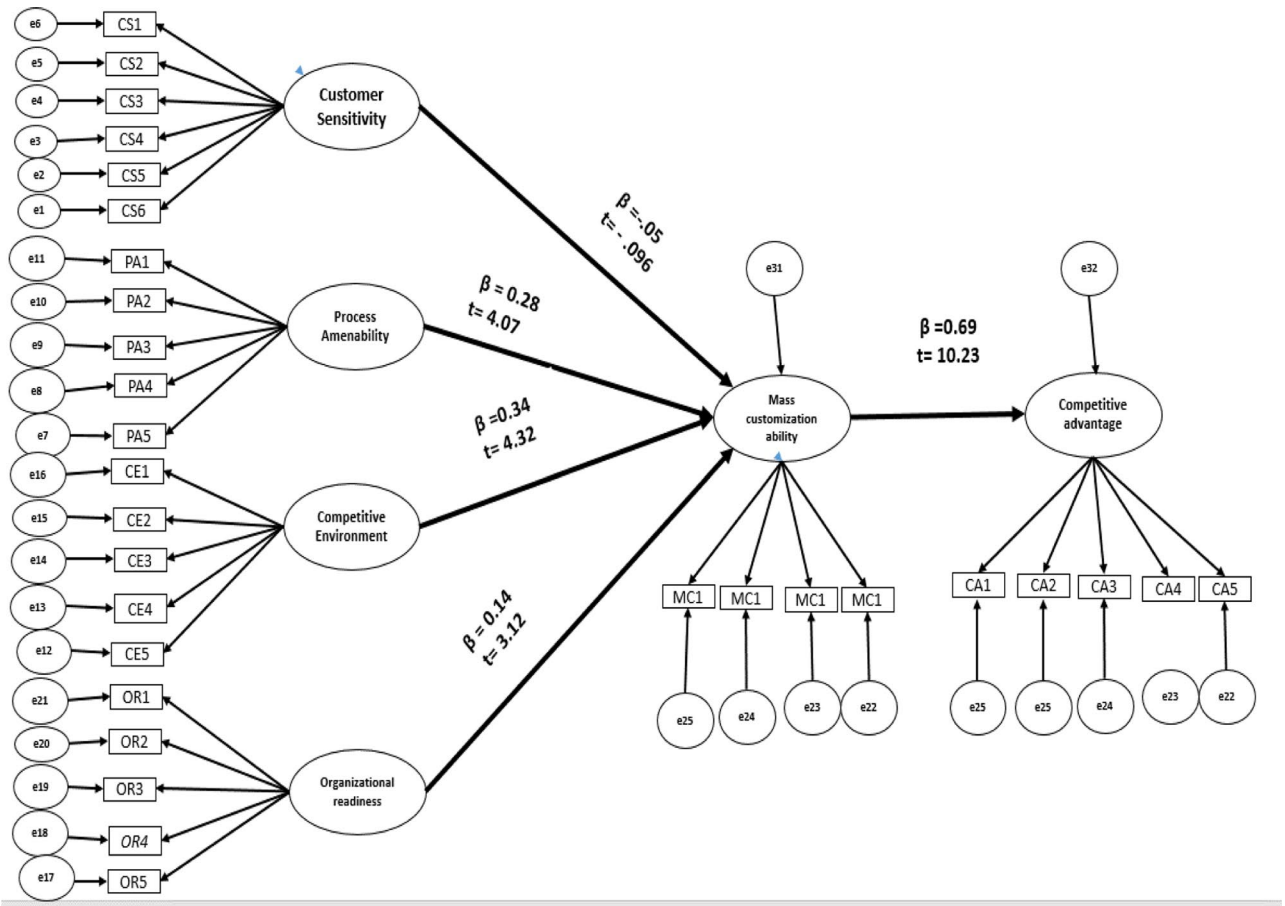


Fig. 3 Path diagram of structural equation model

organizational system transformation, and resource and competency prioritization [86]. Automobile firms of India may expand into MC areas to attract new customers with existing automobiles. They may take use of their strong position in India by following in the footsteps of European and Japanese automakers in providing customized possibilities. Higher prices, sales of add-ons and specific equipment, and a stronger emotional brand loyalty among its customer base could benefit the Indian carmaker [48]. Although such firms have process amenability, organization readiness is essential for MC to be implemented. Another industry with a lot of promise for the MC market is the apparel and footwear business [30]. Because of the presence of a customer-sensitive market, a competitive environment, and organizational readiness owing to a flat organizational structure, MC is viable; nevertheless, process amenability must be strengthened because such companies confront competition both locally and globally.

The disruption caused by COVID-19 on manufacturing has a considerable influence on operational, social, and financial sectors, as well as posing challenges to organizations attempting to accelerate the change of global value distribution models, effectively halting many Indian industries. The manufacturing sector in India is now at a crossroads with unprecedented consequences for manufacturers and supply chains [34]. A new method is required to limit the impact of COVID-19 and rethink risk management and contingency plans. This empirical model of mass customization can provide manufacturers with a solution to the pressing issues that must be addressed to make the business as stable as possible. This strategy will not only improve resilience, protect operations, and support people during the crisis, but it will also help businesses maintain a competitive advantage and accelerate growth after the economy recovers.

6 Managerial implications

The practical goal of the research is to provide strategic instructions to top-level manufacturing executives to encourage mass customization for business excellence. According to the findings of the study, the competitive environment has a significant impact on mass customization, which is consistent with Hart's (1995) research findings that organizations should shift from mass production to mass customization only when market turbulence occurs, resulting in homogeneous versus heterogeneous customer demand. However, the first-mover advantage can only be guaranteed if there is customer loyalty and a good reputation in the market. Because of the turbulence that has disrupted the mass market, mass customization should be regarded as an organizational approach. Firms should view market volatility as an opportunity rather than a hindrance or a threat to which they should escape [1]. For a diverse country like India, organizations need to consider the diverse and individualized needs of consumers, variable human choices, and inconsistent economic level of customers for identifying what to mass customize. An organization should focus on setting up the product vision for the business unit while maintaining a big picture focus; evaluating competitor offerings technically; and commercially and identifying gaps to deliver according to human needs with close control on time to market.

Once the organization had identified its potential to venture into mass customization, it needs to strengthen the other three pillars simultaneously. Customer sensitivity that identifies the uniqueness of customer needs and sacrifices for unmet needs should be a deciding factor for the type of mass-customized product, which should cogitate the voice of the customer during the new design process. The other pillar, process amenability, needs to empower production, process, distribution, the vendor for co-development, and the marketing team to be "market ready" for mass customization. The fourth pillar requires, firstly, top management involvement (highest factor loading) and employee empowerment to ensure a cultural change, followed by skill development of employees for the successful attainment of mass customization goals.

7 Conclusion

With ever-increasing economic instability, MC's ability will inevitably become more crucial for businesses of all sizes to survive and prosper. Given this scenario, businesses preparing to transition from mass production to mass customization will require a framework for execution. While previous research in Indian context aimed at building framework for mass customization implementation, the current research further developed the four pillar of mass customization by eminent researcher Hart [1] and empirically tested the framework identifying the relationships among the constructs of mass customization for competitive advantage. The study was conducted drawing on a sample of 276 Indian companies, using SEM approach to test the developed hypotheses. The results showed that process amenability, competitive environment, and organizational readiness all have a major impact on mass customization. The study, on the other hand, did not support customer sensitivity, which could be because mass customization is a new manufacturing paradigm in India. These scales can be used to enhance prior research that, although focusing on MC ability, failed to take into account the effects of these aspects during the implementation stage.

The above study will aid practitioners, particularly in the Indian setting, by offering a guideline. An empirical

examination of these four pillars can provide a clear indication of how enticing it will be for a company to pursue an explicit mass customization strategy. The significance of a company arriving at mass customization is the culmination of a rational organizational evolution and organizations that already operate in an enlightened realm of operations have the option of mass customization. Organizations need to understand that there is no such thing as a one-size-fits-all company idea when it comes to mass customization and mass customization must be tailored to one's organization's goals, customers, manufacturing capacity, competitive environment, and new technologies accessible at every level. Customer loyalty, market leadership, productivity, and profitability will all be rewarded in abundance for any organization that is truly prepared for the effort of executing this significant new strategic aim demands.

8 Limitations of the study and future research direction

As with any empirical study, some limitations demanded further investigation. This framework should be tested in another country in the future to ensure that it works in other markets. Second, the research concentrates on the manufacturing sector. A cross-industry comparative study could be done in the future to evaluate the model provided in this study in other industries. In the service sector, the approach can also be put to the test.

Future research could include elements not included in this study, such as financial concerns. This framework is best suited for companies with a solid technical foundation. A comparable framework for MSMEs and start-ups may be investigated for the use of mass customization in a variety of businesses. Furthermore, the aforementioned concept was put to the test for mass customization to gain a competitive advantage. It is possible to test the same model to see if it is profitable for other business prospects.

Although mass customization, which employs modularization to improve product variety while maintaining mass production (MP) efficiency, has been popular in recent years, it does have some drawbacks. To begin with, clients do not fully participate in the design process. Second, designers have predetermined possible combinations. Due to multiple challenges faced, partial mass customization can be considered since the scope of operation and organizational readiness required is simpler. Automobile, furniture, and apparel industry can apply partial customization of standard products at the assembly or delivery stages. Case studies in these manufacturing sectors can be further conducted to the study implementation possibilities of MC in India.

Author contribution All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.

Data availability Data will be made available as desired by journal.

Declarations

Ethics approval The authors have abided by the ethics policies of the journal.

Consent to participate Author and co-authors consent to participate

Consent for publication Author and co-authors consent for publication

Conflict of interest The authors declare no competing interests.

References

1. Hart CWL (1995) Mass customization: conceptual underpinnings, opportunities and limits. Int J Serv Ind Manag 6:36–45. https://doi.org/10.1108/09564239510084932
2. Alon I, Pilla V, Bretas G (2021) COVID-19 and international business
3. Deshmukh SG, Haleem A (2020) Framework for manufacturing in post-COVID-19 world order: an Indian perspective. Int J Glob Bus Compet 15:49–60. https://doi.org/10.1007/s42943-020-00009-1
4. Vekic A, Borocki J, Fajsi A (2020) Mass customization strategies in pandemic conditions. 234–240
5. Brown S, Bessant J (2003) The manufacturing strategy-capabilities links in mass customisation and agile manufacturing - an exploratory study. Int J Oper Prod Manag 23:707–730. https://doi.org/10.1108/01443570310481522
6. Liu G, Shah R, Babakus E (2012) When to mass customize: The impact of environmental uncertainty. Decis Sci 43(5):851–887
7. Andújar-Montoya MD, Gilart-Iglesias V, Montoyo A, Marcos-Jorquera D (2015) A construction management framework for mass customisation in traditional construction. Sustain 7:5182–5210. https://doi.org/10.3390/su7055182
8. Kotha S (1996) From mass production to mass customization: the case of the national industrial bicycle company of Japan. Eur Manag J 14:442–450. https://doi.org/10.1016/0263-2373(96)00037-0
9. McCarthy IP (2004) Special issue editorial: the what, why and how of mass customization. Prod Plan Control 15:347–351. https://doi.org/10.1080/0953728042000238854
10. Deshpande A (2017) Article information: relationships between advanced manufacturing technologies, absorptive performance: an empirical investigation. Asia-Pac J Bus Adm. https://doi.org/10.1108/APJBA-03-2017-0024
11. Kristal MM, Huang X, Schroeder RG (2010) The effect of quality management on mass customization capability. Int J Oper Prod Manag. https://doi.org/10.1108/01443571011075047
12. Ahmad S, Schroeder RG, Mallick DN (2010) The relationship among modularity, functional coordination, and mass customization: implications for competitiveness. Eur J Innov Manag. https://doi.org/10.1108/14601061011013221
13. Liu GJ, Shah R, Schroeder RG (2012) The relationships among functional integration, mass customisation, and firm performance. Int J Prod Res 50:677–690. https://doi.org/10.1080/00207543.2010.537390

14. Peng DX, Liu GJ, Heim GR (2011) Impacts of information technology on mass customization capability of manufacturing plants. Int J Oper Prod Manag. https://doi.org/10.1108/01443571111182173/full/html
15. Liu G, Shah R, Schroeder RG (2006) Linking work design to mass customization: a sociotechnical systems perspective. Decis Sci. https://doi.org/10.1111/j.1540-5414.2006.00137.x
16. Wong H, Eyers D (2011) An analytical framework for evaluating the value of enhanced customisation: an integrated operations-marketing perspective. Int J Prod Res 49:5779–5800. https://doi.org/10.1080/00207543.2010.519738
17. Lai F, Zhang M, Lee DMS, Zhao X (2012) The impact of supply chain integration on mass customization capability: an extended resource-based view. IEEE Trans Eng Manag 59:443–456. https://doi.org/10.1109/TEM.2012.2189009
18. Aihua E, Li X, Lu JN (2016) Effects of organizational learning on process technology and operations performance in mass customizers. Int J Prod Econ 174:68–75. https://doi.org/10.1016/j.ijpe.2016.01.019
19. Huang X, Kristal MM, Schroeder RG (2010) The impact of organizational structure on mass customization capability: a contingency view. Prod Oper Manag 19:515–530. https://doi.org/10.1111/j.1937-5956.2009.01117.x
20. Dash P (2019) Mass customization. Int J Psychosoc Rehabil. https://doi.org/10.37200/IJPR/V23I6/PR190782
21. Zhang M, Guo H, Huo B et al (2019) Linking supply chain quality integration with mass customization and product modularity. Int J Prod Econ. https://doi.org/10.1016/j.ijpe.2017.01.011
22. Piller FT, Moeslein K, Stotko CM (2004) Does mass customization pay? An economic approach to evaluate customer integration. Prod Plan Control 15:435–444. https://doi.org/10.1080/0953728042000238773
23. Wong H, Lesmono D (2013) On the evaluation of product customization strategies in a vertically differentiated market. Int J Prod Econ 144:105–117. https://doi.org/10.1016/j.ijpe.2013.01.023
24. Wang Y, Ma HS, Yang JH, Wang KS (2017) Industry 4.0: a way from mass customization to mass personalization production. Adv Manuf. https://doi.org/10.1007/s40436-017-0204-7
25. Zhong RY, Dai QY, Qu T et al (2013) RFID-enabled real-time manufacturing execution system for mass-customization production. Robot Comput Integr Manuf. https://doi.org/10.1016/j.rcim.2012.08.001
26. Shukla M, Todorov I, Kapletia D (2018) Application of additive manufacturing for mass customisation: understanding the interaction of critical barriers. Prod Plan Control 29:814–825. https://doi.org/10.1080/09537287.2018.1474395
27. Yao S, Han X, Yang Y et al (2007) Computer-aided manufacturing planning for mass customization: Part 1, framework. Int J Adv Manuf Technol 32:194–204. https://doi.org/10.1007/s00170-005-0327-z
28. An W, Yang W, Guo W, Zhu D (2014) Research on enterprise customization diagnosis for mass customization. Int J Adv Manuf Technol 76:669–674. https://doi.org/10.1007/s00170-014-6315-4
29. Ullah I, Narain R (2018) Analysis of interactions among the enablers of mass customization: an interpretive structural modelling approach. J Model Manag 13:626–645. https://doi.org/10.1108/JM2-04-2017-0048
30. Purohit JK, Mittal ML, Mittal S, Sharma MK (2016) Interpretive structural modeling-based framework for mass customisation enablers: an Indian footwear case. Prod Plan Control 27:774–786. https://doi.org/10.1080/09537287.2016.1166275
31. Liao K, Deng X, Marsillac E (2013) Factors that influence Chinese automotive suppliers' mass customization capabilities. Int J Prod Econ 146:25–36. https://doi.org/10.1016/j.ijpe.2013.01.014
32. Pedzik M, Bednarz J, Kwidzinski Z et al (2020) The idea of mass customization in the door industry using the example of the company porta KMI Poland. Sustain. https://doi.org/10.3390/su12093788
33. Li DC, Chang FM, Chang SC (2010) The relationship between affecting factors and mass-customisation level: the case of a pigment company in Taiwan. Int J Prod Res 48:5385–5395. https://doi.org/10.1080/00207540903130884
34. Jain P, Garg S, Kansal G (2021) A TISM approach for the analysis of enablers in implementing mass customization in Indian manufacturing units. Prod Plan Control 1–16. https://doi.org/10.1080/09537287.2021.1900616
35. Pirola F, Cimini C, Pinto R (2020) Digital readiness assessment of Italian SMEs: a case-study research. J Manuf Technol Manag 31:1045–1083. https://doi.org/10.1108/JMTM-09-2018-0305
36. Liu GJ, Zhang W, Guo C (2018) Impacts of supply chain planning and integration on mass customization. J Manuf Technol Manag 29:608–628. https://doi.org/10.1108/JMTM-08-2017-0162
37. Tookanlou PB, Wong H (2020) Determining the optimal customization levels, lead times, and inventory positioning in vertical product differentiation. Int J Prod Econ. https://doi.org/10.1016/j.ijpe.2019.08.014
38. Trentin A, Forza C, Perin E (2012) Organisation design strategies for mass customisation: an information-processing-view perspective. Int J Prod Res 50:3860–3877. https://doi.org/10.1080/00207543.2011.597790
39. Kotha S, Yoo J, Park M et al (2004) The manufacturing strategy-capabilities links in mass customisation and agile manufacturing - an exploratory study. Int J Prod Econ 15:707–730. https://doi.org/10.1108/09564239510084932
40. Piller FT (2004) Mass customization: reflections on the state of the concept. Int J Flex Manuf Syst 16(4):313–334. https://doi.org/10.1007/s10696-005-5170-x
41. Zhang M, Lettice F, Zhao X (2015) The impact of social capital on mass customisation and product innovation capabilities. Int J Prod Res 53:5251–5264. https://doi.org/10.1080/00207543.2015.1015753
42. Mourtzis D (2016) Challenges and future perspectives for the life cycle of manufacturing networks in the mass customisation era. Logist Res 9:1–20. https://doi.org/10.1007/s12159-015-0129-0
43. Tang M, Qi Y, Zhang M (2017) Impact of product modularity on mass customization capability: an exploratory study of contextual factors. Int J Inf Technol Decis Mak 16:939–959. https://doi.org/10.1142/S0219622017410012
44. Gunasekaran A, Yusuf YY, Adeleye EO, Papadopoulos T (2018) Agile manufacturing practices: the role of big data and business analytics with multiple case studies. Int J Prod Res 56:385–397. https://doi.org/10.1080/00207543.2017.1395488
45. Dean PR, Tu YL, Xue D (2009) An information system for one-of-a-kind production. Int J Prod Res 47:1071–1087. https://doi.org/10.1080/00207540701543593
46. Zipkin P (2001) The limits of mass customization. MIT Sloan Manag Rev 42(3):8
47. Forza C, Salvador F (2008) Application support to product variety management. Int J Prod Res. https://doi.org/10.1080/00207540600818278
48. Kortmann S, Gelhard C, Zimmermann C, Piller FT (2014) Linking strategic flexibility and operational efficiency: the mediating role of ambidextrous operational capabilities. J Oper Manag 32:475–490. https://doi.org/10.1016/j.jom.2014.09.007
49. Romano P, Vinelli A (2001) Quality management in a supply chain perspective. Int J Oper Prod Manag 21:446–460. https://doi.org/10.1108/01443570110381363

50. Yeung JHY, Selen W, Zhang M, Huo B (2009) The effects of trust and coercive power on supplier integration. Int J Prod Econ 120:66–78. https://doi.org/10.1016/j.ijpe.2008.07.014
51. Ulrich K (1995) The role of product architecture in the manufacturing firm. Res Policy 24:419–440. https://doi.org/10.1016/0048-7333(94)00775-3
52. Fisher M, Ramdas K, Ulrich K (1999) Component sharing in the management of product variety: a study of automotive braking systems. Manag Sci 45:297–315. https://doi.org/10.1287/mnsc.45.3.297
53. Salvador F, Rungtusanatham M, Forza C (2004) Supply-chain configurations for mass customization. Prod Plan Control 15:381–397. https://doi.org/10.1080/0953728042000238818
54. Kotha S (1995) Mass customization: implementing the emerging paradigm for competitive advantage. Strateg Manag J 16:21–42. https://doi.org/10.1002/smj.4250160916
55. Helms MM, Ahmadi M, Jih WJK, Ettkin LP (2008) Technologies in support of mass customization strategy: exploring the linkages between e-commerce and knowledge management. Comput Ind 59:351–363. https://doi.org/10.1016/j.compind.2007.09.003
56. McIntosh RI, Matthews J, Mullineux G, Medland AJ (2010) Late customisation: issues of mass customisation in the food industry. Int J Prod Res 48:1557–1574. https://doi.org/10.1080/00207540802577938
57. Sheng H, Feng T, Chen L et al (2020) Motives and performance outcomes of mass customization capability: evidence from Chinese manufacturers. J Manuf Technol Manag. https://doi.org/10.1108/JMTM-02-2020-0065
58. Tung A, Baird K, Schoch HP (2011) Factors influencing the effectiveness of performance measurement systems. Int J Oper Prod Manag 31:1287–1310. https://doi.org/10.1108/01443571111187457
59. Godinic DB, Du J, Tsoy D et al (2020) Sustaining enterprise operations and productivity during the COVID-19 pandemic: "enterprise effectiveness and sustainability model." Sustain 12:1–27. https://doi.org/10.3390/su12155981
60. Dean PR, Xue D, Tu YL (2009) Prediction of manufacturing resource requirements from customer demands in mass-customisation production. Int J Prod Res 47:1245–1268. https://doi.org/10.1080/00207540701557197
61. Wang Z, Zhang M, Sun H, Zhu G (2016) Effects of standardization and innovation on mass customization: an empirical investigation. Technovation 48–49:79–86. https://doi.org/10.1016/j.technovation.2016.01.003
62. Liu X, Zhao X, Zhao H (2018) Absorptive capacity and business performance: the mediating effects of innovation and mass customization. Ind Manag Data Syst 118:1787–1803. https://doi.org/10.1108/IMDS-09-2017-0416
63. Ellena T, Mustafa H, Subic A, Pang TY (2018) A design framework for the mass customisation of custom-fit bicycle helmet models. Int J Ind Ergon 64:122–133. https://doi.org/10.1016/j.ergon.2018.01.005
64. Malhotra NK, Mukhopadhyay S, Liu X, Dash S (2012) One, few or many? An integrated framework for identifying the items in measurement scales. Int J Mark Res 54:835–862. https://doi.org/10.2501/IJMR-54-6-835-862
65. Das D (2018) Sustainable supply chain management in Indian organisations: an empirical investigation. Int J Prod Res 56:5776–5794. https://doi.org/10.1080/00207543.2017.1421326
66. De Lange P, Jackling B, Suwardy T (2015) Continuing professional development in the accounting profession: practices and perceptions from the Asia Pacific region. Account Educ 24:41–56. https://doi.org/10.1080/09639284.2014.1002800
67. Mendes L, MacHado J (2015) Employees skills, manufacturing flexibility and performance: a structural equation modelling applied to the automotive industry. Int J Prod Res 53:4087–4101. https://doi.org/10.1080/00207543.2014.993772
68. Maklan S, Klaus P (2011) Customer experience: are we measuring the right things? Int J Mark Res 53:5. https://doi.org/10.2501/ijmr-53-6-771-792
69. Reynolds N, Diamantopoulos A (1998) The effect of pretest method on error detection rates. Eur J Mark 32:480–498. https://doi.org/10.1108/03090569810216091
70. Valase K, Raut DN (2019) Mediation analysis of multiple constructs in the relationship between manufacturing and technology and environmental constructs in structural equation model for sustainable manufacturing. Int J Adv Manuf Technol 101:1887–1901. https://doi.org/10.1007/s00170-018-2979-5
71. Schreiber JB, Stage FK, King J et al (2006) Reporting structural equation modeling and confirmatory factor analysis results: a review. J Educ Res 99:323–338. https://doi.org/10.3200/JOER.99.6.323-338
72. Malhotra MK, Grover V (1998) An assessment of survey research in POM: from constructs to theory. J Oper Manag 16:407–425. https://doi.org/10.1016/s0272-6963(98)00021-7
73. Krause DR, Pagell M, Curkovic S (2001) Toward a measure of competitive priorities for purchasing. J Oper Manag 19:497–512. https://doi.org/10.1016/S0272-6963(01)00047-X
74. Podsakoff PM, MacKenzie SB, Lee JY, Podsakoff N (2003) Common method biases in behavioral research: a critical review of the literature and recommended remedies. J Appl Psychol 88:879
75. Fornell CG, Larcker DF (1981) Evaluating structural equation-models with unobservable variables and measurement error. J Mark Res 18(1):39–50
76. Nunnally JC, Bernstein I (1994) Psychometric theory. New York McGraw-Hill
77. Ul Hadia N, Abdullah N, Sentosa I (2016) An easy approach to exploratory factor analysis: marketing perspective. J Educ Soc Res 6:215–223. https://doi.org/10.5901/jesr.2016.v6n1p215
78. Chen IJ, Paulraj A (2004) Towards a theory of supply chain management: the constructs and measurements. J Oper Manag 22:119–150. https://doi.org/10.1016/j.jom.2003.12.007
79. Khurana S, Haleem A, Mannan B (2019) Determinants for integration of sustainability with innovation for Indian manufacturing enterprises: empirical evidence in MSMEs. J Clean Prod 229:374–386. https://doi.org/10.1016/j.jclepro.2019.04.022
80. Shashi CP, Cerchione R, Singh R (2019) The impact of leanness and innovativeness on environmental and financial performance: insights from Indian SMEs. Int J Prod Econ 212:111–124. https://doi.org/10.1016/j.ijpe.2019.02.011
81. Hvam L (2006) Mass customisation in the electronics industry: based on modular products and product configuration. Int J Mass Cust 1:410. https://doi.org/10.1504/ijmassc.2006.010442
82. de Bellis E, Hildebrand C, Ito K et al (2019) Personalizing the customization experience: a matching theory of mass customization interfaces and cultural information processing. J Mark Res 56:1050–1065. https://doi.org/10.1177/0022243719867698
83. Yeung HT, Choi TM (2011) Mass customisation in the Hong Kong apparel industry. Prod Plan Control 22:298–307. https://doi.org/10.1080/09537287.2010.498609
84. Liu N, Chow PS, Zhao H (2020) Challenges and critical successful factors for apparel mass customization operations: recent development and case study. Springer, US
85. Piller FT, Lindgens E, Steiner F (2012) Mass customization at Adidas: three strategic capabilities to implement mass customization. SSRN Electron J. https://doi.org/10.2139/ssrn.1994981
86. Dangayach GS, Deshmukh SG (2001) Implementation of manufacturing strategy: a select study of Indian process

companies. Prod Plan Control 12:89–105. https://doi.org/10.1080/09537280150204022

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

Multimedia Tools and Applications
https://doi.org/10.1007/s11042-022-12854-8

Integrating node centralities, similarity measures, and machine learning classifiers for link prediction

Sameer Anand[1] · Rahul[2] · Abhishek Mallik[2] · Sanjay Kumar[2]

Received: 29 July 2021 / Revised: 21 January 2022 / Accepted: 9 March 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Link prediction is a widely studied topic in graph data analytics and finds numerous applications like friend recommendations in social networks and product recommendations in e-commerce. It refers to predicting new connections or edges that may arise in the near future amongst the nodes of the network. There exist various methods of link prediction generally based on either local, semi-local, or global features of networks and usually suffers from the problems of consistency in their performances over different and large size networks. In this paper, we intend to propose a generic and improved method of link prediction named as NSMLLP by integrating Node centralities, Similarity measures, and Machine Learning classifiers. We calculate popularity measures for every node and evaluate similarity measures for every pair of nodes in the network. The combined popularity and similarity measures form the features for every node pair in the network. The combined features of the nodes at the end of the edges, along with the positive or negative edge label, form a well-defined dataset for the task of link prediction. This dataset is then fed into machine learning classifiers like Random Forest classifier, AdaBoost classifier, and an ANN based classifier. The results obtained from these classifiers are then combined to make the final link prediction. We provide an information gain study aiming to quantify the improvement brought on by our proposed method. A feature importance study is also provided to comprehend better the relative importance of the various popularity and similarity measures used by us. The experimental results obtained on multiple real-life networks demonstrate that the proposed technique outperforms many existing popular methods of link predication based on several evaluation criteria.

Keywords Artificial neural network (ANN) · Complex networks · Link prediction · Node centrality · Random forest · Similarity measures

✉ Sanjay Kumar
sanjay.kumar@dtu.ac.in

Extended author information available on the last page of the article.

1 Introduction

The study of complex networks provides an enormous scope of data analytics and renders a wide range of research problems. The various real-life networks like wireless communication networks, transportation networks, web networks, power networks, biological networks, and social networks have become an integral part of our daily lives [17, 51]. These are complex networks containing a large number of nodes having intricate relationships among them. These networks are dynamic and grow over time. Complex network analysis deals with mining and analyzing the voluminous graph data and predicting valuable patterns, which can be useful for numerous real-life applications like influence maximization, information propagation, community detection [33, 34, 36]. Link prediction is one of the popular research topics in the field of complex network analysis. Link prediction refers to the process of analyzing the network based on various parameters to predict the possibility of new edges that may appear in the near future [26]. The concept of link prediction has a very long history in the study of complex networks. A variety of researchers from different backgrounds, ranging from biology to computer science, have approached this concept since its advent [41]. Ever since the idea of predicting links was developed, it has become a hot issue in complex network analysis. Friend recommendation in online social network is a popular application of link prediction which deals with prediction of new edges or connection between unconnected users based on features, hobbies, and interests of the users. Some other notable applications of link predictions in real-life scenario are predicting disease outbreaks, planning transportation, product recommendation in e-commerce applications, hotel recommendation, music recommendation and many others [1, 28, 32, 52, 66].

Link prediction helps us to apprehend the growth of a network as most real-life networks, like online social networks, communication networks grow over time. Hence, link prediction can very well be associated with information and knowledge engineering in a connected world. The process of link prediction is based on two things, the popularity of a node pair and the similarity of a node pair. This relation seems to be logical as a popular node is more likely to connect with other nodes, and a pair of nodes that are similar enough have more chances of being connected in the future. Most of the existing methods for performing link prediction considered local information or single similarity index or paths of a single snapshot of the network, which leads to less accurate predictions in complex networks. These techniques can be broadly categorised into similarity based, maximum likelihood based, probabilistic based, random walks based, and matrix factorization based methods [26, 41]. Some recent work has also been done by exploring the prospects of machine learning and deep learning techniques [27, 39, 57].

In this paper, we exploit the possibility of a probable link amongst two nodes depending on the possible influence of one node over the other, i.e., the popularity of the node, along with the similarity amongst two nodes. We introduce an algorithm by integrating **N**ode centrality, **S**imilarity measures and **M**achine **L**earning classifier for **L**ink **P**rediction named as **NSMLLP**. The proposed method is based on the fusion of four node centralities or popularity measures (i.e., Extended Coreness, PageRank, Gateway Local Rank, and Local Informational Dimensionality index) and five similarity measures (namely, Common Neighbors, Adamic Adar, Cosine Similarity, Matrix Forest Index, and path-based LHN2). Usually, the existing link prediction algorithms capture only a particular aspect of the network while ignoring the rest of the information. These methods are generally based on local, semi local or global features of the network and suffers from performance degradation on different networks or when the size of the network increases. This is basically due to the fact

that a single type of method may not perform well on all kinds of graphs. Also, as the size increases, the feature extraction capability of more global methods decrease while the more local methods require better integration for proper link prediction. Our purpose of choosing the aforementioned centrality and similarity measures is to better capture and combine the local, quasi-local and global features of nodes in the network. The calculated node centralities and similarity measures form the features for every pair of nodes in the network. The node pair forms a edge which may or may not be present in the network. The edges are labeled as positive or negative on the basis of their occurrence in the network. These edge features along with their corresponding labels forms a well-labeled dataset. The well-labeled and well-balanced dataset is then fed into machine learning classifiers like, Random Forest classifier, AdaBoost classifier and artificial neural network based classifier. The results obtained from these classifiers are then combined to make the final link prediction in the network. To validate the efficacy of the proposed algorithm, we compare its performance with various existing popular methods on eight different datasets based on the evaluation criteria like AUC, precision, recall, and F1 score. In summary, the main contributions of the paper are as follows.

- We propose a generic link prediction method by integrating node centralities, similarity measures, and various machine learning classifiers, thereby obtaining a weighted combinatorial link prediction algorithm.
- We present an algorithm to convert a network into a well-labeled and well-balanced dataset suitable for machine learning and deep learning classifiers.
- We present an information gain study to reveal the improvement in various performance metrics attained by our proposed NSMLLP algorithm.
- We also provide a feature importance study to understand better what type of node centrality and similarity measures tend to be more salient towards link prediction.
- Our proposed method produces commendable results compared to many popular and recent methods of link prediction based on simulation performed on eight real-life networks using various performance metrics.

The rest of the paper is structured as follows. In Section 2 some of the existing work done in the field of link prediction is discussed. Section 3 illustrates the preliminaries required to gain a better understanding of the proposed work. The proposed algorithm is formulated and discussed in detail in Section 4. Section 5 describes the datasets and evaluation metrics used by us for our study. The experimental results are discussed in Section 6. Finally, Section 7 presents the concluding remarks and the scope for future work.

2 Related work

Link prediction tends to be one of the most explored topics in network analysis because of its numerous applications, due to which a lot of ideas have been proposed for improving the obtained solution to this problem. Numerous link prediction techniques have been presented in the literature. These techniques can majorly categorised into similarity based, maximum likelihood based, probabilistic based, random walks based , and matrix factorization based methods [44]. Liu and Lü [40] introduced a node similarity index based on local random walk (LRW) for link prediction with lower computational complexity as compared to other

random-walk based similarity indices. The LRWLP method computes the similarity score between two nodes v_i and v_j as follows:

$$Score(v_i, v_j) = \frac{k_i \times k_j}{|E|^2} \tag{1}$$

where k_i, k_j, $|E|$ denote the degree of node v_i, degree of node v_j, and the total number of edges in the network, respectively. Monteserin et al. [43] proposed an influence maximization-based approach to predict the links of target users to influential users with one limitation that the target must have influential users in their current neighborhood to be able to receive recommendations of other influential links. Biswas et al. [16] proposed a community based link prediction method in their work. They considered the effect of communities in link prediction. Edge centrality measures were used by them to consider the likelihood of missing links. Das [20] introduced a positive and negative link prediction technique in social networks using the notion of sentiment analysis. The method worked by analyzing the users posting the reply of posts with positive or negative emotions, which can further help in examining the learning ability of the user and utilize it for recommending the people with positive or negative emotions. Zhou et al. [67] proposed the H-Salton as an improved alternate to Salton (also called Cosine Similarity) index for link prediction. The H-Salton replaces the degree with h-index. The method computes the similarity score between two nodes v_i and v_j as follows.

$$Score(v_i, v_j) = \frac{|\Gamma(v_i) \cap \Gamma(v_j)|}{\sqrt{h(v_i) \times h(v_j)}} \tag{2}$$

where $h(v)$ denote the value of h-index for the node v. A low-rank matrix completion algorithm was suggested by Gao et al. [23]. Their algorithm used the proximal forward-backward splitting to minimize the nuclear norm simultaneously. Wang et al. [55] fused several symmetric and asymmetric matrices and adjacency matrix to achieve a fusion model based framework for link prediction. Tripathi et al. [53] proposed a hybrid approach combining the local as well as global features of the networks to make link prediction. Mistele et al. [42] employed a neural network based algorithm to predict a particular author's citation count and h-indices in a citation network. Kumar et al. [35] used a level-2 node common neighbors based clustering information and generated a similarity score based on the retrieved information. By considering the importance of social balance and status theories as much as possible, Gu et al. [25] proposed a link prediction algorithm on signed social networks based on latent space mapping. In their study, Kim et al. [29] highlighted the formational constraints that need to be considered to increase the pragmatic value of link prediction algorithms targeted to achieve success for collaboration networks. A deep convolution neural network based algorithm for link prediction was proposed by Wang et al. [58]. They experimented with a densenet and resnet based architectures. Yadav et al. [63] also presented a community structure based link prediction algorithm. Their design included a feature-based similarity measure that incorporates the community structure along with other network features for link prediction. Aziz et al. [10] aimed at improving the existing path based link prediction algorithms by incorporating the node information along the path. Kaya [28] proposed hotel recommendation framework based on the idea of link prediction which relies on user's location. First, they created a customer-hotel bipartite network and then the relationship information of this network was used. Then via supervised learning link prediction is done based on the user's location. Zhang et al. [66] presented a new music recommendation approach based on the concept of link prediction utilizing bipartite graph homogeneity. They incorporated the complex representation-based link prediction

(CORLP) method along with attribute extraction and similarity extraction. Wu et al. [61] in their works proposed an influential node identification technique to formulate a node-ranking based link prediction metric. Ashraf et al. [54] proposed Newton's gravitational law-inspired strategy to combine various indexes to improve their overall link prediction capabilities. Based on that literature, we took the best combination they suggested, that is, the CCM (closeness centrality and MFI) combination, which is defined as:

$$Score(v_i, v_j) = CC(v_i) \times CC(v_j) \times (MFI(v_i, v_j))^2 \tag{3}$$

where $CC(v)$ denotes the value of Closeness Centrality for the node v and $MFI(v_i, v_j)$ denotes the MFI score for the node pair (v_i, v_j). Wu et al. [62] proposed a node ranking based approach of link prediction for both scale-free time-evolving networks and real-life dynamic networks. Li et al. [39] proposed an ensemble-based link prediction algorithm named EMLP that combines various similarity measures, which are further passed onto an ensemble machine learning model comprising of logistic regression and XGBoost classifier. However, the EMLP approach has not utilized the importance of node centrality, and also, the method lacks in the study on information gain and feature importance. Wang et al. [56] proposed a multidimensional link prediction algorithm and evaluated the effects of different dimensions on the prediction of user's links. Bahera et al. [13] proposed a link prediction method named LPXGB using the XGBoost machine learning technique. They utilized different graph feature techniques to represent the dataset suitable for binary classification problems for performing link prediction. Ghorbanzadeh et al. [24] proposed a hybrid common neighborhood based link prediction algorithm in consideration of the fact that even if there is no common neighbor among two nodes, they still have a chance of probable link in the future. It also predicts the direction of the predicted link for the network. Recently, various link prediction methods based on matrix factorization, network embedding methods including random walks and deep walks [14, 15, 22, 45] were introduced. Random walk is one of the primary similarity-based methods used for link prediction, which detects similarities between nodes by randomly going through the graph in global and quasi-local ways. Some link prediction methods also leveraged the spatio-temporal correlations [4, 8] between the nodes of the network for prediction of links and various flows in the network. Some modified versions of the long short-term memory (LSTM) and convolutional neural network (CNN) were used to explore temporal relations and spatial relations of the network. Another domain of link prediction is spatiotemporal data-based traffic flow prediction. One of the recent advancements in this field is the data aggregation based DSHTNet model [7]. It used parametrically combines several features like closeness volume, daily volume, trend volume, and the external branch to predict the traffic flows in every region of the city with very increased accuracy. In extension to DSHTNet, the authors also proposed an attention-based AAtt-DHSTNet to predict short-term traffic flows. The algorithm was tested on both online and offline scenarios using edge or fog infrastructure. Several other approaches have also been proposed in this field using various hybrid-neural networks and attention based neural networks [5, 6]. Recently, a Graph Convolutional Network (GCN) based approach for predicting traffic flows was also presented [9]. They also present an attention based approach to make short-term traffic flow predictions. The approaches were also tested on online and offline frameworks using edge-fog architectures.

Over the years, a lot of work has been done in the field of link prediction. In most of the existing literature, the features and topological details are extracted from the entire network. Usually, the entire network dataset is split into training and testing networks by hiding some connections which are to be predicted while testing the model. But the features are calculated for the entire network, which is not meaningful as the connections which are to be

predicted cannot be used for estimating the features on which the model is to be trained. On the other hand, we calculate the features from the training network while training the model. Through this work, we aim at improving this prediction quality by suitably combining various classifiers. Most of the existing works on link prediction use either global, quasi-local, or local features of the network. Such an approach leaves out several necessary topological features of the network. All these points motivate us to propose a novel link prediction algorithm that uses several node centralities and similarity measures along with various machine learning classifiers.

3 Preliminaries

This section provides the background information utilized in this work. Here, we briefly present the various node centralities and the similarity measures used. It also describes the random forest classifier employed by us to make link predictions finally.

3.1 Node centrality

Node centrality-based methods are the most common way to measure the popularity of the nodes in the network. These methods assign a score to nodes based on the importance of their topological position under various considerations in the network [30, 31]. We utilize the four different node centralities namely, extended coreness (EC) centrality [11], PageRank (PR) [46], gateway local rank (GLR) [49], and local informational dimensionality (LID) [59] in our study. The extended coreness centrality and PageRank (PR) are classical node centralities. However, gateway local rank (GLR) and local informational dimensionality are recently introduced popular methods to measure the popularity of nodes for their spreading capabilities. We utilize these four different node centralities to adequately capture the local, quasi-local, and global features of nodes in the network. A brief description of each adopted centrality is as follows.

1. Extended Coreness (EC) Centrality: Based on the notion that nodes lying in the core of the system are vital for spreading influence, Kitsak et al. [65] introduced k-shell centrality. The method prunes the nodes from the external to the inner shell and assigns k-shell value to each node level-wise. One primary deficiency of this method is that all the nodes in the same shell are expected to have equal spreading capacity. To address this, Bae et al. [11] proposed extended coreness centrality of a node by adding the k-shell value of its neighbors, allowing to distinguish the spreading ability of the nodes having the same k-shell value. Mathematically extended coreness (EC) centrality of a node is presented as follows:

$$EC(v) = \sum_{u \in \Gamma(v)} Ks(u) \tag{4}$$

   where $Ks(v)$, $EC(v)$, $\Gamma(v)$ represent k-shell value, extended coreness value of a node v, and set of neighbor nodes of v respectively.
2. PageRank (PR): PageRank is based on the random walk in the network [46]. The Google search engine uses the idea of PageRank centrality to rank the web pages based on their relevance. The random walk theory assumes an imaginary surfer who is randomly clicking on links. Hence, it outputs a probability distribution representing the

likelihood that a person randomly clicking on links will arrive at any particular page. PageRank is defined as:

$$PR(v) = (1 - d) + d * \sum_{u \in \Gamma^{-}(v)} \frac{PR(u)}{deg(u)} \tag{5}$$

In (5), d is the damping factor, deg gives the value of degree of the given node, and Γ^{-} gives the set of nodes which are Inbound to the given node.

3. Gateway Local Rank (GLR): The gateway local rank [49] is based on the local structure of nodes and uses the local properties of the nodes to locally detect the critical nodes over the graph. It first split the graph into various communities and then find local critical and gateway nodes in that community. The final influencing score of the node is calculated as the sum of shortest paths to those local critical and gateway nodes.
4. Local Informational Dimensionality (LID): The local informational dimensionality [59] index is based on the local information dimensionality and states that the nodes with a higher local information dimensionality are more influential in the complex network, where the Shannon entropy is used to consider the information about the nodes.

3.2 Similarity measures

Similarity measures based methods for link predictions are intuitive, in which for each pair of nodes v_i and v_j, a similarity score $S(v_i, v_j)$ is calculated. The value of score between each pair of nodes that are not yet connected can be computed using different structural and topological network properties. The high value of similarity measures can correspond to the predicted link between them.The following five similarity measures are used in our study.

1. Common Neighbors (CN): This is a local similarity metric which is based on the number of mutual friends of the given node pair. Common Neighbor [64] index is defined as:

$$CN(v_i, v_j) = \Gamma(v_i) \cap \Gamma(v_j) \tag{6}$$

In (6), Γ is the set of neighbors of the given node.

2. Adamic Adar (AA): It is an extended local similarity measure which depends on the number of mutual neighbors relative to the nodes' degrees. Adamic Adar [3] is defined as:

$$AA(v_i, v_j) = \sum_{v_k \in \Gamma(v_i) \cap \Gamma(v_j)} \frac{1}{\log \Gamma(v_k)} \tag{7}$$

where Γ denotes the set of neighbors of the given node.

3. Cosine Similarity (COS+): Cosine Similarity or Salton [50] is a similarity index that measures the similarity between two vectors of an inner product space. In networks theory, this is estimated as the cosine of the angle between the vector representations of the two nodes. Cosine Similarity is defined as:

$$COS(v_i, v_j) = \frac{\Gamma(v_i) \cap \Gamma(v_j)}{\sqrt{deg(v_i) \times deg(v_j)}} \tag{8}$$

In (8), Γ denotes the set of neighbors of the given node. deg gives the value of degree of the given node.

4. Leicht–Holme–Newman 2 (LHN2): The LHN2 [37] is a path based similarity index which states that, if the neighbors of node x are similar to node y, then nodes x and

y are similar. That is, the similarity of the nodes is transitive. LHN2 index is defined as:

$$LHN2 = 2M\lambda_1 D^{-1}\left(I - \frac{\phi}{\lambda_1}A\right)^{-1} D^{-1} \tag{9}$$

where D is the degree matrix, ϕ is an adjust parameter with value $0 < \phi < 1$. λ_1 is the largest eigenvalue of the adjacent matrix A. M is the total number of network edges.

5. Matrix Forest Index (MFI): This is a similarity index which is based on the matrix forest theory in data structures. The matrix forest index [19] for a node pair is defined as the ratio of the number of spanning rooted forests, such that the two nodes belong to the same tree to all spanning rooted forests in the entire network. Matrix forest index is defined as:

$$MFI = (I + L)^{-1} \tag{10}$$

where L is the Laplacian Matrix of the network.

3.3 Machine learning models

The following are the machine learning models used in our algorithm.

1. Random Forest Classifier: Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges to a limit as the number of trees in the forest becomes large [18]. It is one of the most successful ensemble learning techniques which have been proven to be very popular and powerful techniques in the pattern recognition and machine learning for high-dimensional classification and skewed problems.
2. AdaBoost Classifier: The AdaBoost algorithm is one of the most popular machine learning algorithms. The AdaBoost algorithm select various weak classifiers iteratively from a number of candidate weak classifiers and combine them linearly to form a strong and more powerful classifier. The algorithm also corrects the misclassifications made by weak classifiers as it select more and more weak classifiers. AdaBoost is less susceptible to overfitting than most learning algorithms and is very fast if simple weak classifiers like decision trees are chosen.
3. Artificial Neural Network: An Artificial Neural Network consists of layers of neurons where the input of neurons in one layer is the output of the neurons in the previous layer. It generally consists of an input layer, one or more hidden layers of neurons and a final layer of output neurons. The output, h_i, of neuron i in the the hidden layer is given by:

$$h_i = \sigma\left(\sum_{j=1}^{N} W_{ij}x_j\right) \tag{11}$$

In (11), $\sigma()$ is the activation function, N is the number of neurons in the previous layer, W_{ij} are the weights and x_j are the outputs by previous layer or inputs to the current layer.

4 Proposed methodology

This section prsents our proposed algorithm by integrating node centrilities, similarity measures and machine learning classifiers for link prediction. The proposed NSMLLP algorithm

incorporates four different node centralities, five different similarity measures, and three machine learning classifiers, namely, random forest classifier, AdaBoost classifier, and artificial neural network based classifier. Hence, we intend to introduce a generic framework for link prediction by combining the various node features and classifiers. The set of the node centralities and similarity measures are chosen so as to better capture the local, semi-local and global properties of the nodes in the network. Since various centrality and similarity measures have strong correlations, over-fitting can occur quite easily if we directly apply machine learning algorithms to them. Hence, we select a definite set of popularity and similarity indexes representing the nodes' various information and are not minutely correlated. We combine the results of three machine learning classifiers by taking simple average of prediction produced by them. For improving the performance of our algorithm, we also used various techniques like cross-validation and grid searching to determine the best hyper-parameters of all the three models. Based on this investigation, we finally arrived at the values of hyper-parameters. This paper considers the problem of link prediction in complex networks as a binary classification problem in which the main task is to predict the existence and non-existence of a link between a pair of nodes based on the existing network structure. For this task, the nodes' similarity and centrality indexes are considered the features for the machine learning model, and the model itself combines these indexes to make the predictions. The existence and non-existence of the links amongst two nodes form the positive or negative label, respectively. We preprocess the raw data to obtain proper training and testing datasets for our model. We start by splitting the entire raw data into a training and testing set based on a splitting ratio varied from 50% to 95%. Then we generate labels for both the sets based on the existence and non-existence of edges amongst two nodes where the various centrality and similarity measures of the nodes form the features for them. The flow diagram of the proposed algorithm is shown in Fig. 1. It shows that how in step 1, we first split the entire network into a training, G_T and testing network G'. Then in step 2, we balance the number of existent and non-existent edges in both the networks by including an equal number of non-existent edges as there are existent edges. Then in step 3, we go on to create a training dataset, D_T and testing dataset D' from G_T and G' respectively. The dataset is created by generating features for the edges by computing various node centralities and similarity measures. The labels for the edges are generated on the basis of existence and non-existence of the edges in the network. In step 4, we train all the three classifiers chosen by us for making the link predictions. Finally in step 5, we combine the results obtained by all the classifiers to make the final link predictions.

The various algorithmic steps involved in our proposed method are described below.

4.1 Creation of training and testing graphs G_T and G'

Suppose a complex network is represented as $G(V, E)$ where V is the set of vertices in the graph and E is the set of edges in the graph. Now, we create a new graph $G_T(V, E_T)$ where V is the same set of vertices as in the original graph G and E_T is the subset of original link set E, produced by sampling the links randomly. The size of the edge set E_T obtained after sampling of the original graph depends on the splitting ratio in which the entire graph is split. This splitting ratio is varied from 50% to 95% in steps of 5% for the experiments performed. This reduced graph $G_T(V, E_T)$ is used for training our machine learning model for link prediction. This trained classifier is further used to make predictions of the linkages amongst the nodes which belong to the edge set $E - E_T$. Hence, the problem of link prediction reduces to the prediction of graph $G(V, E')$, where $E' = E - E_T$, based on the information from the graph $G_T(V, E_T)$. This reduction has been performed to make

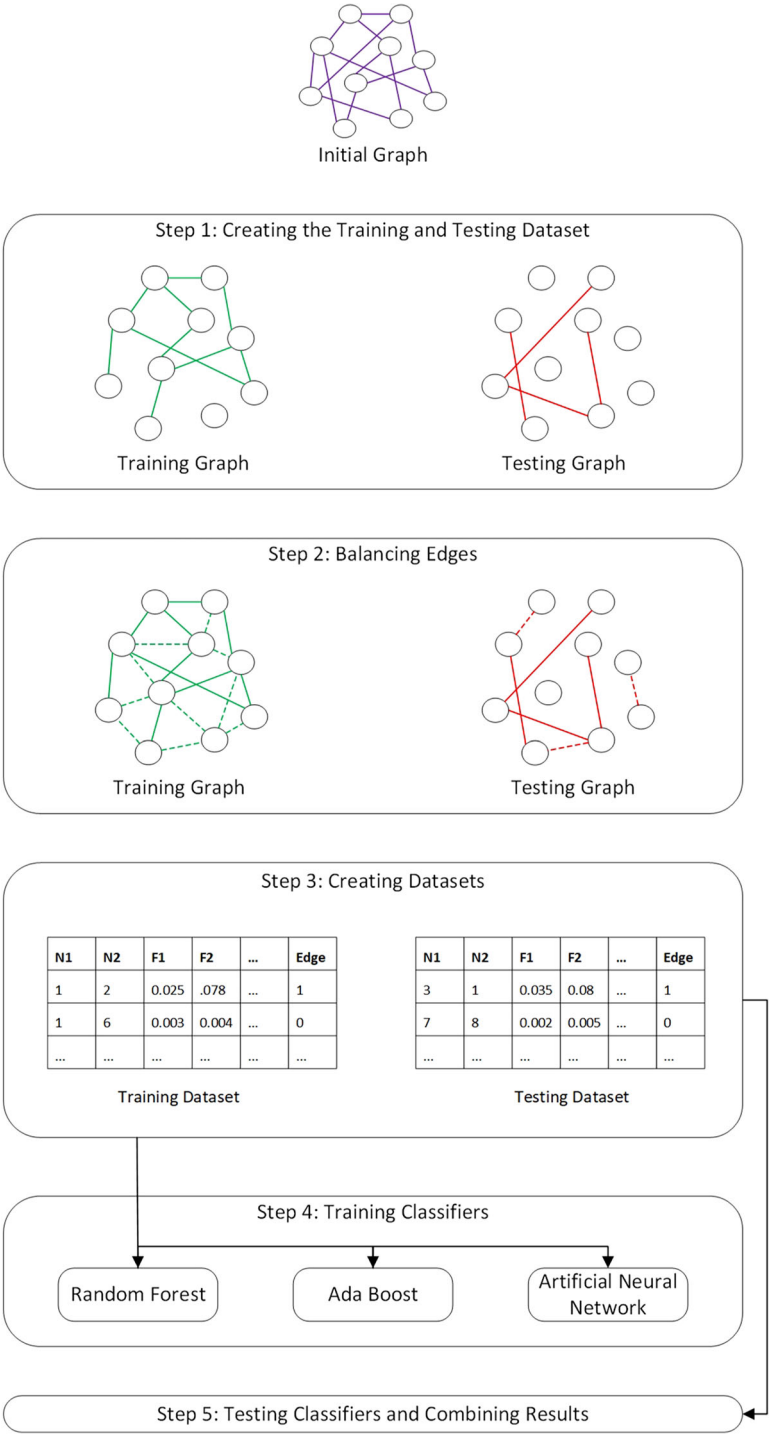


Fig. 1 Flow Diagram of the proposed algorithm

the situation similar to a real-life scenario where we predict links from the current graph and not using the edges which will be present in the graph in future. This comes as an improvement to the various existing link prediction algorithms which train the classifiers based on the entire network, which is not the case in a realistic setting. So finally we have a training graph, G_T and testing graph, G' as follows:

$$G_T = G(V, E_T),\ 0.5 \leq \frac{\mid E_T \mid}{\mid E \mid} \leq 0.95 \tag{12}$$

$$G' = G(V, E - E_T) \tag{13}$$

4.2 Balancing the type of edges

Usually, in complex networks, the edge matrix is very sparse, i.e., the number of connected node pairs is much lower than the node pairs which are not connected. This sparseness is even more pronounced in social networks in particular. Hence, the dataset that we utilize by considering all existing and non-existent edges may be highly unbalanced. If we apply machine learning algorithms to this unbalanced data, they are more likely to over-fit the dataset for negative samples, i.e., non-existent edges, and render the predictions meaningless. Thus, in this step, we try to balance the dataset by using the down sampling strategy. In down sampling, we select all pair of nodes which are connected. We chose all the links in set E_T and E' and then randomly extract the same number of node pairs that are not connected by edges. These form new sets of complement edges $\overline{E_T}$ and $\overline{E'}$.

Let $\mid V \mid = n$ be the number of vertices in the network and $edge(u, v)$ represent an edge between node u and v. Then let the set of all possible edges in the network be E^U defined as follows:

$$E^U = \{edge(u, v), u, v \in V \ and \ \mid E^U \mid = \frac{n * (n-1)}{2}\} \tag{14}$$

Also, we define the complement of E_T as $\overline{E_T}$ and the complement of E' as $\overline{E'}$ which represents the set of non-existent edges in the original network. These are represented as follows:

$$\overline{E_T} = \{edge(u, v), u, v \in \{E^U - E\} \ and \ \mid \overline{E_T} \mid = \mid E_T \mid\} \tag{15}$$

$$\overline{E'} = \{edge(u, v), u, v \in \{E^U - E\} \ and \ \mid \overline{E'} \mid = \mid E' \mid\} \tag{16}$$

Thus G_T and G' are now represented as follows:

$$G_T = G(V, \{E_T \cup \overline{E_T}\}) \tag{17}$$

$$G' = G(V, \{E' \cup \overline{E'}\}) \tag{18}$$

4.3 Construction the well-labeled dataset

This step discusses the procedure of creating the well-labeled dataset D_T and D' for training and testing purposes from the G_T and G' respectively. These datasets contain the features for edges as well as their corresponding labels in the network. For creating the features of the network, we chose extended coreness (EC), PageRank (PR), gateway local rank (GLR), and local Informational dimensionality (LID), common neighbors (CN), Adamic Adar (AA), cosine similarity (COS+), matrix forest index (MFI), and path-based LHN. Figure 1 is a pictorial representation of our proposed algorithm. Here, $Node$ represents the souce and target nodes. While $F1$, $F2$ represent the calculated features like GLR, LID etc. These features are represented in Section 4.3. The edge represents the existence and non-existence of

the links. The purpose of choosing the varied set of node features is that all these centralities capture different network features. For instance, LID and Extended Coreness consider local information about each node of a network. GLR is based on the global knowledge of each node as distributed by the communities. Page rank is a global popularity measure based on the random walk process. Common neighbor index and Adamic Adar are the local similarity measures that consider only some node pair neighbors. Cosine similarity is based on the random walk process between the node pair. LHN2 is based on the similarity path between the pair of nodes. Matrix forest index is derived from the matrix-forest theory. By combining all these indices, we make use of all the local, quasi-local and global information provided by these centralities. Let $p(v)$ be the function that calculates all the popularity measures for the node v. Let $s(u, v)$ be the function that calculates the similarity measure amongst the nodes u and v. Using these functions we obtain a set of popularity as well as similarity measures for the nodes of the network as follows:

$$P = \{p(v), v \in V\} \tag{19}$$

$$S = \{s(u, v), u, v \in \{E_T \cup \overline{E_T}\}\} \tag{20}$$

$$S' = \{s(u, v), u, v \in \{E' \cup \overline{E'}\}\} \tag{21}$$

Here, P is the set of popularity measures for all the nodes in the network while S and S' is the set of similarity measures for the edges in training and testing networks. These form the features for our dataset. These features were calculated by considering the topological details of only the training network, G_T as in a realistic scenario, we only have the training network for feature extraction. The dataset also a contains a *label* attribute representing the existence of edges in the network. Let $L(u, v)$ be the function which generates the *label* for the edge represented by the nodes u and v. If there is a connection, the value of label is set as 1. Otherwise, the value of label is set as 0. More formally it can be represented as:

$$L(u, v) = \begin{cases} 1, (u, v) \in E \\ 0, (u, v) \in E^U - E \end{cases} \tag{22}$$

Finally, we obtain a well-labeled and well-balanced training and testing dataset, D_T and D', respectively. These datasets can more formally be represented as follows:

$$D_T = \{P(u) \cup P(v) \cup S(u, v) \cup L(u, v), (u, v) \in \{E_T \cup \overline{E_T}\}\} \tag{23}$$

$$D' = \{P(u) \cup P(v) \cup S'(u, v) \cup L(u, v), (u, v) \in \{E' \cup \overline{E'}\}\} \tag{24}$$

4.4 Training the classifiers

In last step, we generated well-labeled and well-balanced datasets D_T and D' for training and testing our classifiers. Now, the training and testing of our proposed algorithm take place on this dataset. For any machine learning based classification task, we require a training set on which to train the classifier and the testing set on which to test the prediction capability of the classifier. Following along these lines, we divide our dataset D^{new} into training as well as testing dataset according to a particular splitting ratio. As part of our proposed algorithm, we use three machine learning models to predict the links, thus, in this step all three models are trained:

1. Random Forest Classifier: The Random Forest is chosen as one of the classifiers as it is an ensemble based model that combines a bunch of weak learners, tree-based classifiers to arrive at a better prediction capability. The hyper-parameters used for the model are:

estimators as 100, with maximum depth per tree being six and splitting criterion as "Gini" index. Let the trained Random Forest model be represent as:

$$RF = RandomForest(D_T) \tag{25}$$

2. AdaBoost Classifier: The Ada-Boost Classifier is used as this instead of processing the decision trees in parallel, it uses a sequential approach and keeps on learning from its mistakes. This has been used alongside of Random Forest Classifier as they both use different approaches for approaching the same problem, thus we will get better and more generalized results. The hyper-parameters used for the model are: estimators as 200, and the weak classifier being used is Decision Tree Classifier with maximum depth per tree being one. Let the trained Ada-Boost model be represent as:

$$Ada = AdaBoost(D_T) \tag{26}$$

3. Artificial Neural Network: The ANN is chosen an an classifier as most of the networks are very complex in nature and the ANN is pretty good at learning complex features about the data. The hyper-parameters used for the model are: epochs as 100, with a batch size of 64 and the loss function being "Binary Cross-entropy" and "Adam" as the optimizer. Let the trained ANN model be represent as:

$$ANN = ArtificialNeuralNetwork(D_T) \tag{27}$$

4.5 Combining the classifiers and making predictions

After training all the classifiers, we go on to making predictions for the testing dataset. The individual predictions made by each classifier is combined by taking their weighted mean. For a classifier C, the weight of its prediction towards the final result is decided by using the wt_C function described as below:

$$wt_C = \frac{\Sigma_{i=1}^{i=len(D')} \left\lfloor \frac{C(D'[i]-L[i])+L[i]}{2} \right\rfloor}{\Sigma_{i=1}^{i=len(D')} C(D'[i]-L[i])} \tag{28}$$

The wt_C function has been designed to give more importance to the classifier making the correct predictions towards the existence of the links that actually exist. The weighted mean method is better than just averaging the outcomes of each model thereby having an equal contribution to the final prediction. This can be understood as some models are known to perform much better or much worse than other models and averaging them on the same scale might give less optimal results. On the other hand, we use a weighted average which helps in giving appropriate importance to an algorithm based on its standalone performance. Here, $len(D')$ represents the length of the dataset D' while $L[i]$ is the label for the i^{th} entry in the dataset or the ground truth for the existence of the link in the network. So, $C(D'[i]-L[i])$ represents the predictions made by the classifier on the dataset D' by excluding the label attribute, $L[i]$ for the i^{th} entry in the dataset. Let this be represented as $T[i]$. The results from all the classifiers are combined as follows:

$$\Theta[i] = \frac{RF(T[i]) * wt_RF + Ada(T[i]) * wt_Ada + ANN(T[i]) * wt_ANN}{wt_RF + wt_Ada + wt_ANN} \tag{29}$$

In this way, the prediction for $T[i]$ is made according to the following equation:

$$Prediction[i] = \begin{cases} 1, & \Theta[i] \geq 0.5 \\ 0, & \Theta[i] < 0.5 \end{cases} \tag{30}$$

Hence, the final results for our proposed link prediction algorithm are calculated as follows:

$$Results = \{Prediction[i],\ i = 1, 2, ..., len(D')\} \tag{31}$$

5 Datasets and evaluation metrics

This section describes the various datasets and the evaluation metrics used by us for the purpose of this study:

5.1 Datasets

The following data sets were used to verify the effectiveness of the proposed NSMLLP algorithm:

1. American Aviation Network (USAir) [12]: It consists of 332 airports and 2126 routes, where each node in the network corresponds to an airport, and each edge depicts whether there is a direct flight between two airports or not
2. Nematode Neural Network (CE) [60]: This network represents the neural network of C. Elegans. It consists of 277 nodes and 2105 edges, where each node represents one nematode neuron, and each edge represents neuronal synaptic connections with other neurons
3. Small and Griffith and Descendants Citation Network (SmaGri) [12]: It consists of 1024 nodes and 4913 edges, where each node represents a publication by the SmaGri and each edge represents a citation of SmaGri's publications
4. FB-Food (FB) [48]: The dataset represents the food page network of Facebook where the nodes represent the Facebook food pages and the links represent the likes among them. It consists of 620 nodes and 1483 links.
5. Aves-wildbird (Aves) [21]: It represents a real world animal interaction data set where the nodes represents the wild birds and the links represent the interaction between these wild birds. It consists of 202 nodes and 11900 edges.
6. PolBlogs (PB) [2]: The PolBlogs network represents the interaction between US US political blog pages where the nodes represent blog pages and the edges represent the hyperlinks among them. It consists of 1222 nodes and 19021 links.
7. EmailEU [38]: This network represents the interaction via emails for a large European research institution where the nodes represent the users while the edges represent interactions amongst them. There are 32430 nodes and 54395 edges.
8. Epinions [47]: Epinions network represents the interaction for the customer review platform Epinions where nodes represents users and edges represent trust relationships amongst the users. There are 26588 nodes and 100120 edges in the network.

The statistical information of the various datasets used by us is summarised in Table 1.

5.2 Evaluation metrics

To evaluate and compare the proposed NSMLLP algorithm with other link prediction methods, we used the following performance matrices:

1. AUC ROC: Area under the receiver operating characteristic (ROC) curve (AUC) is one of the popular evaluation metric for link prediction algorithms. AUC can be defined as

Table 1 Statistical descriptions of various datasets used

Dataset	#Nodes	#Edges	Avg. Degree	Density
USAir	332	2126	12	0.038
CE	277	2105	8	0.019
SmaGri	1024	4913	9	0.019
FB	620	1483	6	0.011
Aves	202	11900	117	0.586
PB	1222	19021	7	0.011
EmailEU	32430	54395	3	0.001
Epinions	75879	508837	7	0.001

the probability that a randomly chosen missing link is given a higher score than a randomly chosen non-existent link. After providing a proximity score to each missing and nonexistent link, we randomly selected a missing and nonexistent link and compared their scores to compute the AUC value. If among n independent comparisons, the missing links have higher score n' times and the same or less score n'' times, the AUC value is defined in (32) [67]:

$$AUC = \frac{n' + 0.5n''}{n} \tag{32}$$

2. Precision: The Precision score to evaluate the various algorithms, which is the ratio of correctly predicted links to the total predicted links. The precision score is given as:

$$Precision = \frac{TP}{TP + FP} \tag{33}$$

In (33), TP denotes the correctly predicted links, that is, these predicted links are present in the original network while FP denotes the incorrectly predicted links, which means that, these links are not present in the original network.

3. Recall: The Recall Score is the most commonly used performance measure for the unbalanced data, hence we included this in our study. It is defined as:

$$Recall = \frac{TP}{TP + FN} \tag{34}$$

In (34), TP denotes the correctly predicted links, that is, these predicted links are present in the original network while FN denotes the incorrectly predicted links, which means that, these links are not present in the original network.

4. F1 score: The F1 score is mainly used to find the overall accuracy of the algorithm and is defined as the weighted average of precision and recall as shown in (35):

$$F1\ Score = \frac{2 * Precision * Recall}{Precision + Recall} \tag{35}$$

5. Information Gain: The information gain attained by an algorithm refers to the gain in knowledge by an algorithm as compared to its peer algorithms. Formally, it can be stated as the improvement in the value of a performance metric achieved by an algorithm as compared to its competing algorithms for the underlying task. It can mathematically be expressed as:

$$\text{Information Gain} = \frac{Metric(Proposed\ Algo) - Metric(Best\ Algo)}{Metric(Best\ Algo)} \tag{36}$$

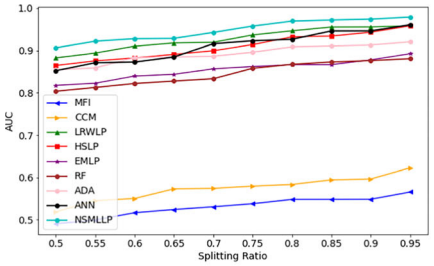
Here, $Metric(Proposed\ Algo)$ is the value of the performance metric obtained by the algorithm while $Metric(Best\ Algo)$ is the value of the performance metric obtained by the algorithm which gives the best results amongst all the considered algorithms.

6 Results

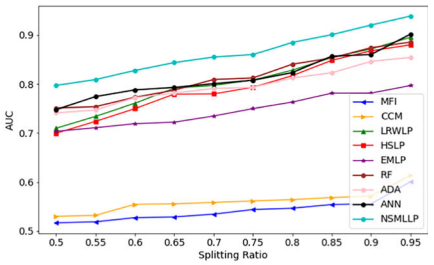
In this section, we present the description of the experimental setup adopted by us, along with a discussion on the performance of our proposed NSMLLP algorithm as pitted against several contemporary and classical methods for link prediction. We performed the investigation on eight datasets of various sizes, nature, and applications and obtained the values of different performance metrics as depicted in Section 5. Since the link set E is randomly sampled to create the data set D, and the selection of a missing link and a non-existent link for comparison is also random, we experimented multiple times for each evaluation metrics. For each data set, all the performance metrics were computed ten times, and results are averaged over. To observe the effect of splitting ratio of link set E, splitting ratio was varied from 50% to 95% with a step of 5%. We also performed an information gain study for our algorithm to analyzed how much improvement is provided by our algorithm in terms of various performance metrics chosen by us. At last, a feature importance study is also presented to understand the major contributing node centralities and similarity measures that help in the task of link prediction. To compare the efficacy of the proposed NSMLLP algorithm, we compared its performance with a total of eight different methods of link prediction where three methods, namely, random forest (RF), AdaBoost (ADA), and artificial neural networks (ANN), are baseline machine learning classifiers and solves link prediction problem as a binary classification problem. Moreover, we have also used the combination of these three classifiers in a weighted manner as a part in our proposed approach. The five other popular and recent algorithms that we used for the comparison are Matrix Forest Index (MFI) [19], local random walk based link prediction (LRWLP) [40], CCM Index [54], H-Salton Index based link prediction (HSLP) [67], and EMLP Algorithm [39]. We have used python programming language and its environment to perform the simulations. All the machine learning and deep learning classifiers, node centralities and similarity measures are implemented in Python using the libraries scikit-learn and networkx respectively. All the experiments are performed on a personal computer with Intel i5-7200U 2.50GHz processor with 8GB of RAM, 1TB HDD and 128GB SSD. Now, we present below the performance comparison of the proposed method with various existing considered methods for link prediction based on each evaluation criteria considered in this study.

6.1 AUC ROC

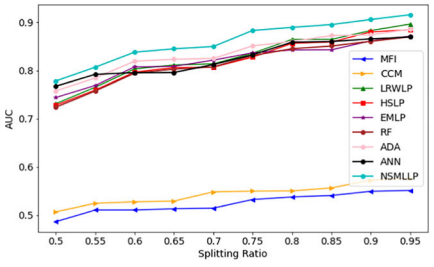
The AUC ROC values obtained on various datasets by various algorithms are shown in Fig. 2. The value on the x-axis represents the splitting ratio for the dataset, whereas the y-axis denotes the obtained AUC value. The splitting ratio varies from 50% to 95%. For example, the value of splitting ratio as 0.95 signifies that training is done on 95% dataset while testing is done on 5%. In the USAir dataset, it can be seen that our algorithm is beating all the other methods, for all the splitting ratio followed by LRWLP, HSLP, and ANN anf remaining other methods. In the CE dataset, the proposed algorithm performs better than all other algorithms and the performance is relatively very good. In the FB-Food data set, the proposed algorithm has outperformed all the other algorithms by a significant margin over all the splitting ratios. However, LRWLP and H-Salton algorithms are not too far away. In



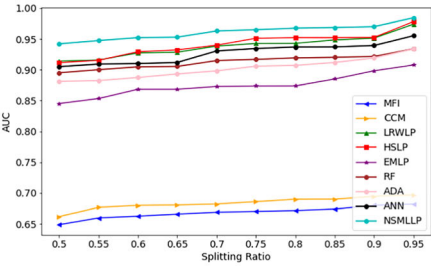
(a) USAir



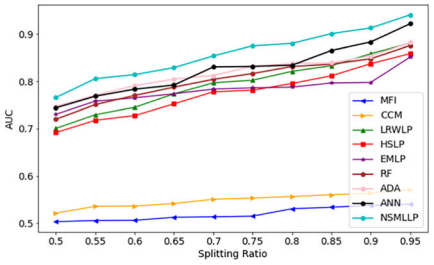
(b) CE



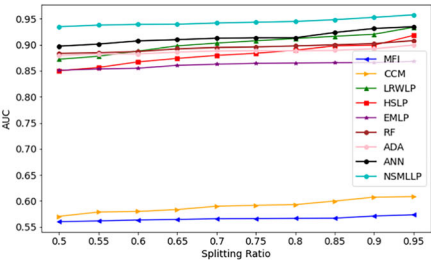
(c) FB-Food



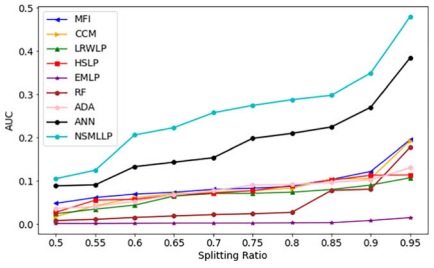
(d) Aves



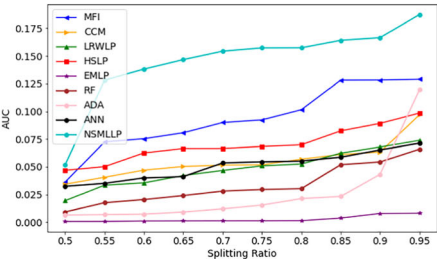
(e) SG



(f) PolBlogs



(g) Email



(h) Epinions

Fig. 2 AUC ROC results obtained on various datasets for different algorithms. The simulations have been executed 100 times, and the obtained result is averaged over them

the Aves dataset, our algorithm is at the top for all splitting ratios. In the SG data set, the proposed algorithm again outperformed all the other algorithms with a significant margin for all splitting ratios. In the PolBlogs data set, our algorithm is at the top for all splitting ratios. Although, the increase in AUC ROC values with respect to the increase in splitting ratio is not as high as compared to other datasets. The NSMLLP performs better than all its individual models on all the datasets signifying a successful combination of models. Thus in terms of the AUC evaluation measure, it can be concluded that the proposed NSMLLP algorithm performed better than all other algorithms in all the datasets for almost all values of the splitting ratio.

6.2 Precision and recall

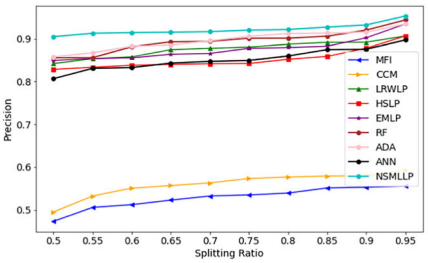
The Precision and Recall scores obtained on the various datasets with different link prediction methods are shown in Figs. 3 and 4 respectively. The value on the x-axis represents the splitting ratio for the dataset, whereas the y-axis denotes the obtained precision or recall value. The splitting ratio varies from 50% to 95%. The trends on both the parameters, namely, precision and recall, are almost similar. In USAir, FB-Food, and PolBlogs datasets, although the proposed method produces the best and consistent result, other methods like HSLP, EMLP, random forest are very close to our method. However, in the case of CE, Aves, SG, Email, and Epinions datasets, our method performs exceedingly well. There is a significant gap between the results of the proposed method and other methods. For SG and Aves datasets, the second-best result is produced by the HSLP method followed by the AdaBoost classifier and EMLP. In the case of the Email dataset, artificial neural network classifier performs best after the proposed method. Similarly, in the Epinopns dataset, the second-best performer after our method is MFI. Hence after observing the results, it is evident that the proposed NSMLLP algorithm performed well and consistently on almost all the datasets for each value of the splitting ratio.

6.3 F1 score

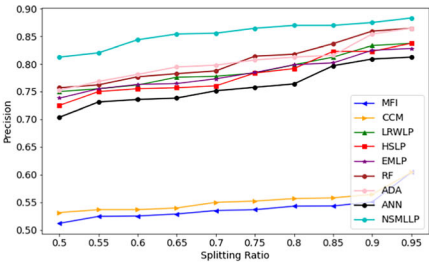
The F1 score result obtained by various algorithms on various datasets are shown in Fig. 5. The value on the x-axis represents the splitting ratio for the dataset, whereas the y-axis denotes the obtained F1 score value. The splitting ratio varies from 50% to 95%. By observing the results, it is evident that the F1 score values scored by the NSMLLP algorithm are overall good. Also, the NSMLLP performs better than all its individual machine learning classifiers, namely, random forest, Adaboost and artificical neural network on all the datasets signifying a successful combination of models. Hence, we can conclude that the proposed method produces improved results and outperforms all other methods considered in comparison in terms of F1 score for all splitting ratios in each dataset.

6.4 Information gain

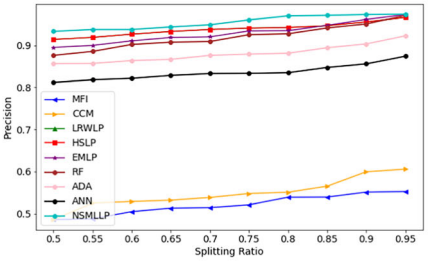
The information gain for an algorithm can be stated as the improvement in the value of the desired performance metric attained by the proposed algorithm when compared against the existing algorithms. As part of our work, we calculate the information gain attained by our algorithm in terms of AUC ROC, precision, recall, and F1 score. Figure 6 shows the information gain for the proposed algorithm over other algorithms considered in comparison based on the four metrics for all the datasets. This shows the percentage improvement that our proposed work obtained in terms of AUC ROC, precision, recall and F1 score with



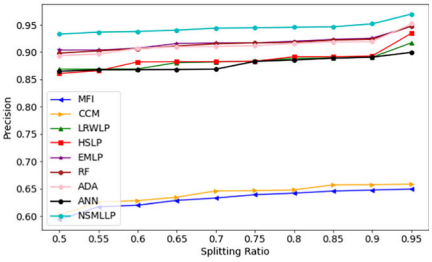
(a) USAir



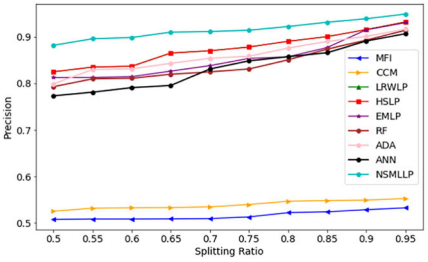
(b) CE



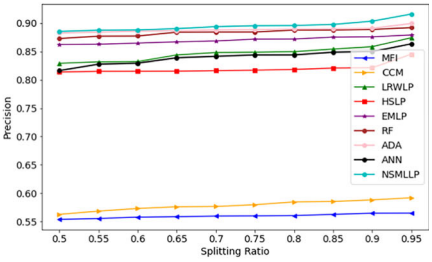
(c) FB-Food



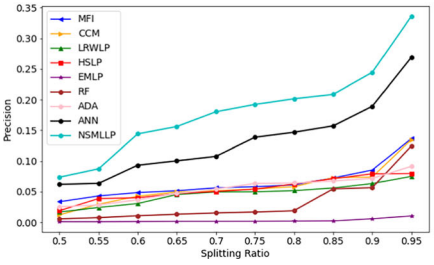
(d) Aves



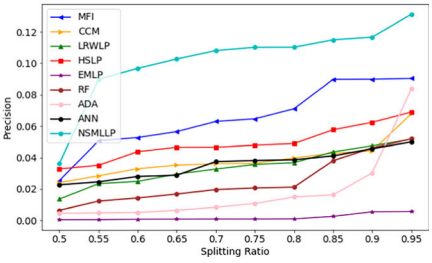
(e) SG



(f) PolBlogs

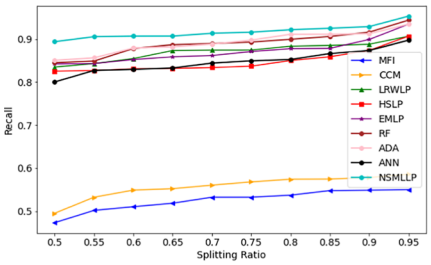


(g) Email

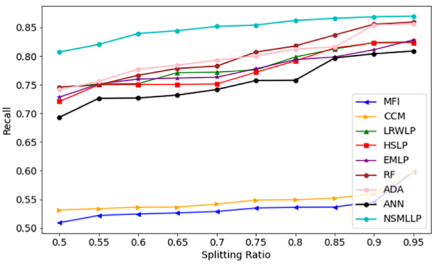


(h) Epinions

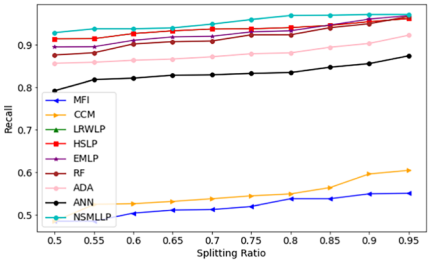
Fig. 3 Precision results obtained on various datasets for different algorithms. The experiments were run 100 times and then their average was taken to achieve the results



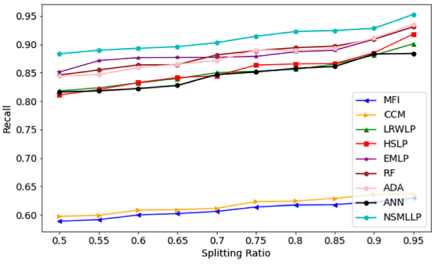
(a) USAir



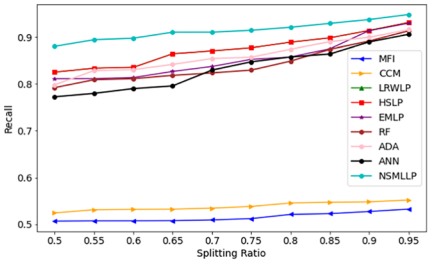
(b) CE



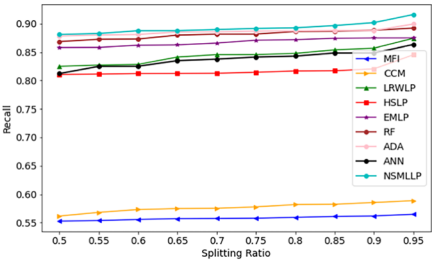
(c) FB-Food



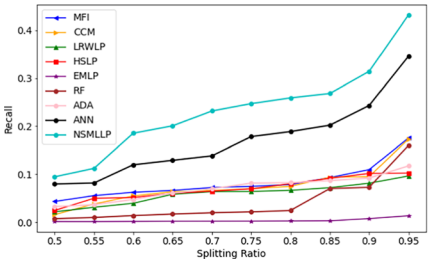
(d) Aves



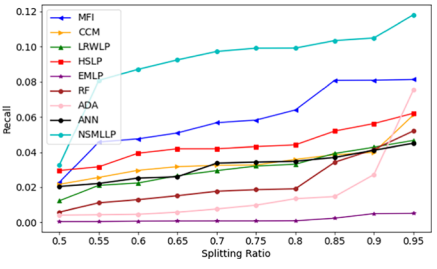
(e) SG



(f) PolBlogs

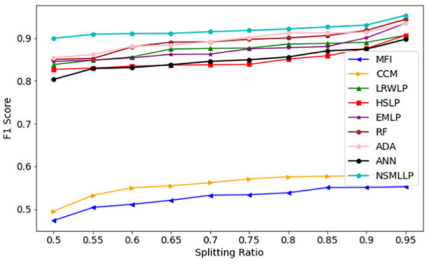


(g) Email

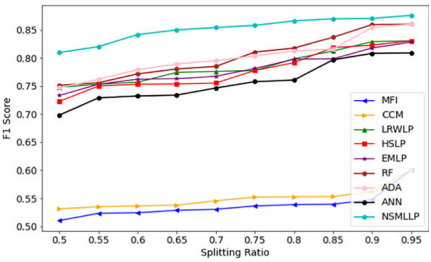


(h) Epinions

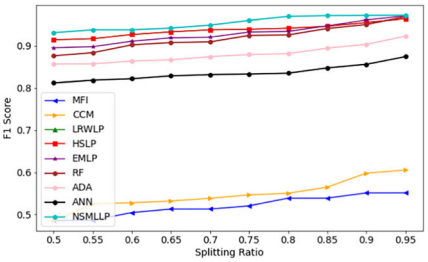
Fig. 4 Recall results obtained on various datasets for different algorithms. The experiments have been executed 100 times, and the obtained result is averaged over them



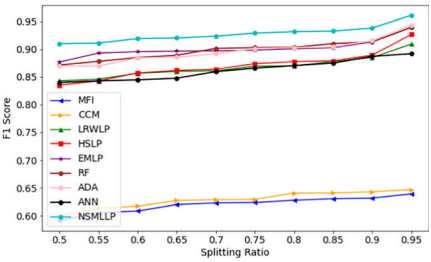
(a) USAir



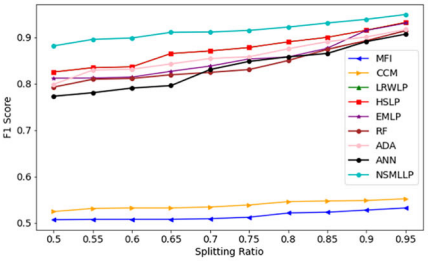
(b) CE



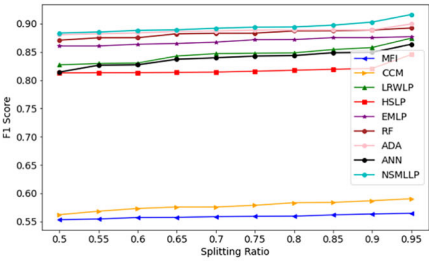
(c) FB-Food



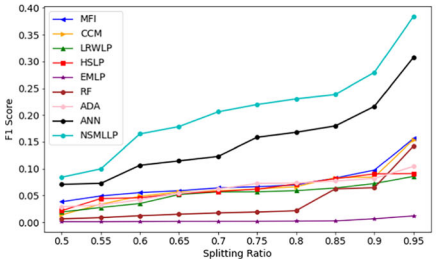
(d) Aves



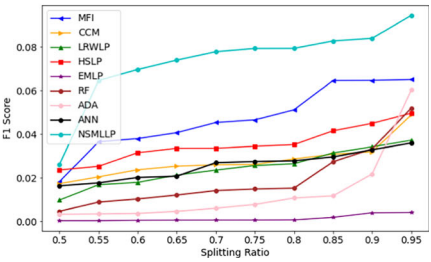
(e) SG



(f) PolBlogs



(g) Email



(h) Epinions

Fig. 5 F1 score results obtained on various datasets for different algorithms. The experiments have been executed 100 times, and the obtained result is averaged over them

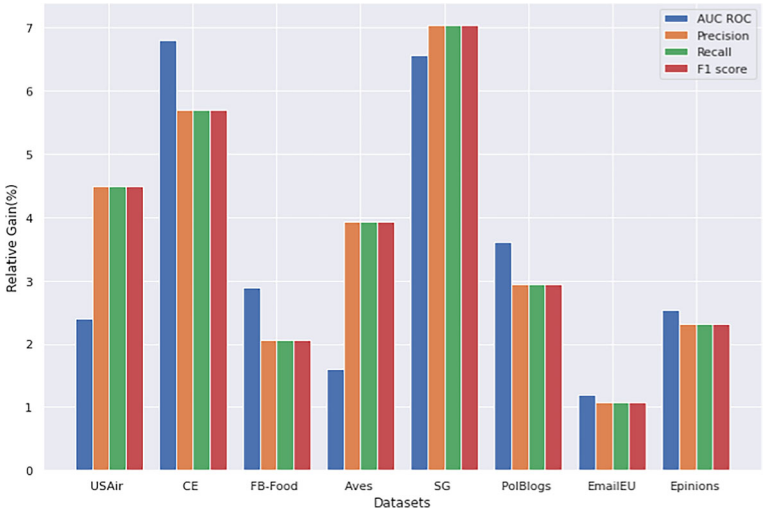
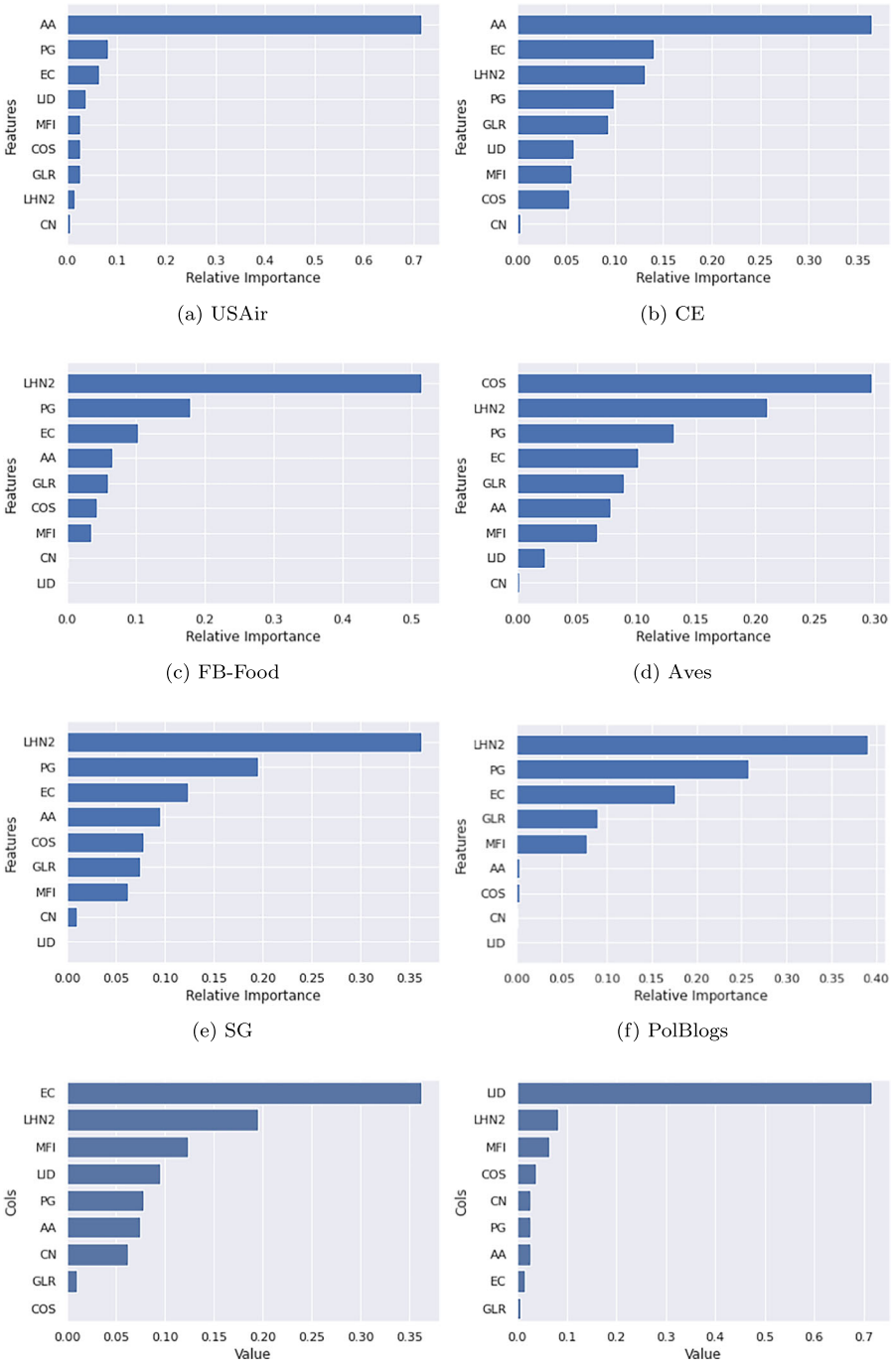


Fig. 6 Relative information gain of proposed algorithm in terms of AUC ROC, precision, recall and F1 score against the chosen algorithms for all the datasets

respect to the best performance obtained by the other algorithms under consideration. For our analysis, we have considered the splitting ratio to be 80%, i.e., for training, we have used 80% of the dataset, while for evaluating the approach, we used 20% of the dataset. A positive value shows that our proposed algorithm improved the value of that metric, while a negative value shows a decrement in the value of that metric as compared to the baseline algorithms. As can be seen from the figure that we obtain only positive values for all the performance metrics for all the datasets. This implicates that our proposed work obtained an improved performance in terms of all the chosen performance metrics as compared to the results obtained by all the other algorithms under consideration. The highest information gain for the AUC metric is achieved for the CE network, while the least is achieved for the Aves network. The gain in precision, recall, and F1 score tends to be the same for all the networks due to the high correlation between these performance metrics. The maximum gain is attained for the SG network, and the minimum gain is attained for the FB-Food network for these performance metrics. The above analysis shows that our proposed method provides improved performance for various real-life networks tested against different contemporary algorithms.

6.5 Feature importance

For our algorithm, we utilized various node centralities and similarity measures as features for the nodes. All these features have varying amounts of contributions in aiding the algorithm towards the task of link prediction. Figure 7 shows the relative feature importance of all the chosen features with the splitting ratio set to 80% for all the chosen networks. For the USAir network, the Adamic-Adar similarity measure tends to be the most important feature. This is further followed by the PageRank and Extended Coreness values, and the rest of the features tend to follow thereby. For the CE network also the Adamic-Adar similarity measure is the most crucial feature as per the NSMLLP algorithm. The rest of the features lag far behind the Adamic-Adar similarity measure, but they are still a substantial contributor



(a) USAir (b) CE

(c) FB-Food (d) Aves

(e) SG (f) PolBlogs

(g) EmailEU (h) Epinions

Fig. 7 Relative feature importance of the chosen centralities and similarity measures over all the datasets for the task of link prediction. The splitting ratio was set to be 80%

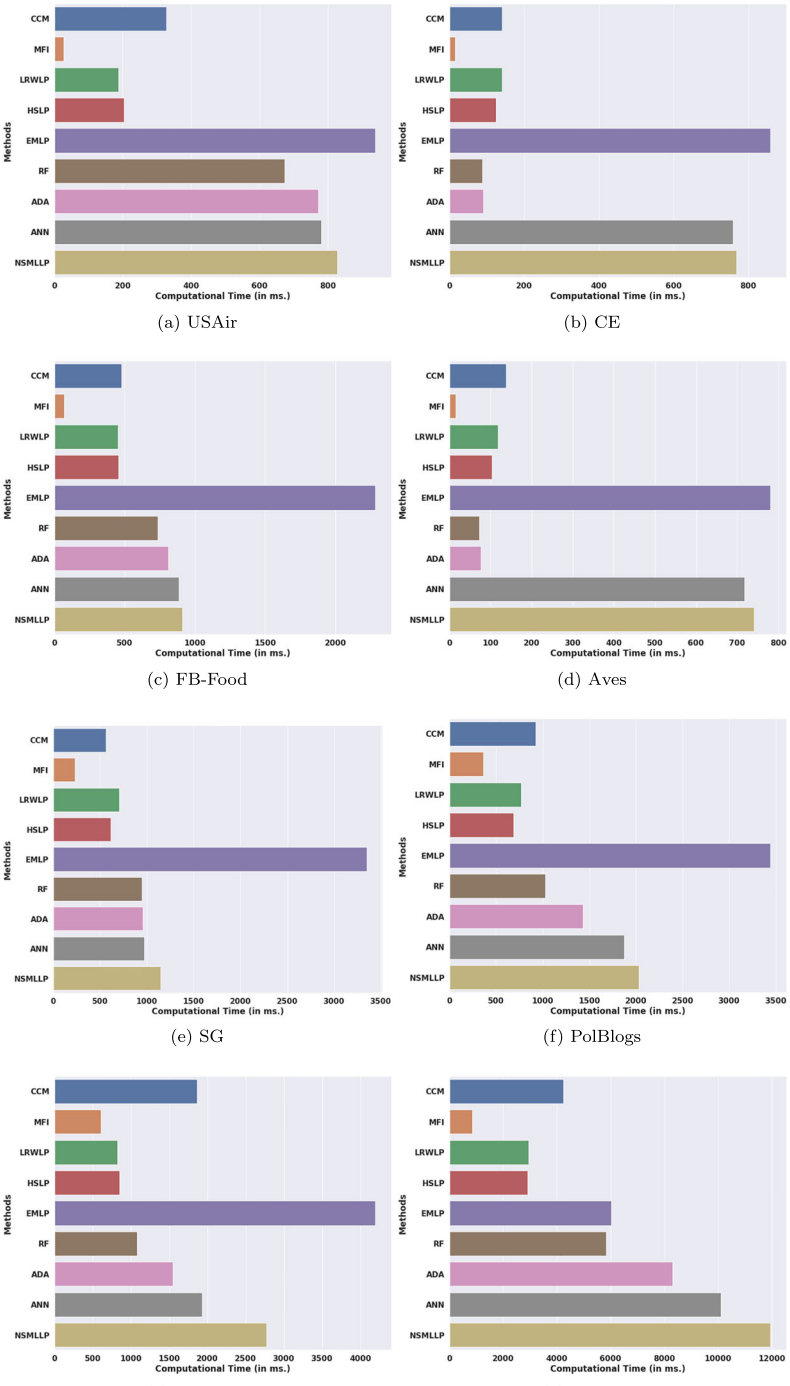
towards the task of link prediction. However, Common Neighbor based similarity measure is the least contributing feature having a feature importance value close to 0. For the FB-Food network, the LHN2 value is the most important feature. All other node centralities and similarity measures tend to follow this. Common Neighbor similarity measure and the LID node centrality measure are the poorest contributors. For the Aves network, the COS similarity measure is the most helpful feature towards link prediction. The rest of the node centralities and similarity measures also contribute substantially to this network. However, the common neighbor based similarity measure still tends to be the least important feature towards link prediction. The LHN2 similarity measure is the most crucial feature for our proposed link prediction algorithm for the SG network. All the other features follow them, thereby making a considerable contribution towards the task of link prediction. The Common Neighbors similarity measure and the LID node centrality still tend to be the poorest contributor for link prediction. For the PolBlogs network, the LHN2 is the most important contributor, closely followed by the PageRank, extended coreness, and other chosen node centralities and similarity measures. However, for this network, even LRWLP and Cosine similarity measures also don't perform well. The Common Neighbors and LID still tend to be the worst contributors towards link prediction. For the EmailEU network, extended coreness centrality is amongst the most important feature followed by LHN2, while COS similarity is the worst performer. On the other hand, we observe that LID is the best performer for the Epinions network while GLR is the worst performer. This can be summarised by the fact that even though being competitive towards the task of link prediction, the LHN2, PageRank, and Extended Coreness values are the most important features for the link prediction. The rest of the features are also substantial contributors to the task of link prediction. The Common Neighbors and LID values tend to be the poorest contributors.

6.6 Computational time analysis

In this section, we compare the computational time complexity, in terms of absolute time, of the various algorithms over all the datasets. For the purpose of experiments, the dataset is split in an 80:20 ratio. Figure 8 represents the time taken in milliseconds (ms) by all the algorithms to perform link prediction over all the datasets. We can see that for all the networks except for EmailEU and Epinions, which are relatively larger datasets, our algorithm performs in under 2 seconds. The computational performance of our proposed NSMLLP is better than the EMLP method, which is based on an ensemble machine learning approach. However, our method requires slightly more absolute time (in ms) than the baseline machine learning classifiers. Our proposed algorithm's high computational time complexity can be attributed to the deep learning architecture that it incorporates in its functioning and the time complexity associated with calculating the various node centralities and the similarity measures.

7 Conclusion

Link prediction refers to predicting edges or links that may arise in the near future based on the available information and features of the nodes in the network. It tends to be one of the widely studied topics in the field of network analysis. In this paper, we introduced an improved method of link prediction named NSMLLP by integrating node centralities, similarity measures, and machine learning classifiers through this work. The node centralities and similarity measures are chosen to cover local, semi-local, and global features of the



(a) USAir (b) CE

(c) FB-Food (d) Aves

(e) SG (f) PolBlogs

(g) EmailEU (h) Epinions

Fig. 8 Computational time (in ms.) taken by all the algorithms over all the datasets for the task of link prediction. The splitting ratio is taken to be 80:20

nodes appropriately. The existence and non-existence of edges form the labels for the edges for every node pair in the network. The node features, along with their labels, are then fed to three machine learning classifiers, namely, random forest, AdaBoost, and artificial neural network based classifier. The results obtained from these classifiers are then combined in a weighted manner to make the link prediction as a binary classification task. Based on cross-validation and grid searching, we determined the various hyper-parameters for our model. We performed the experiments on eight real-life networks and compared results with several contemporary and classical algorithms for link prediction. The obtained results suggest that the proposed NSMLLP algorithm outperforms other methods considered in the comparison. Along with that, we also performed the computational time analysis of the various algorithms over all datasets and found that the proposed algorithm performs the task of link prediction under 2 seconds in most of the datasets which is at par with the performance of other ML and deep learning based algorithms. An information gain study is also presented in this work to capture the improvement attained by our proposed work. We also presented a relative feature importance study to understand better what type of node centralities and similarity measures are the most important for link prediction. The results obtained by the proposed algorithm based on the various evaluation metrics, information gain, and the feature importance study advocate the supremacy of our algorithm and validate our choice of the node centralities, similarity measures, and machine learning classifiers.

Declarations

Conflict of Interests The authors declare that they have no conflict of interest.

References

1. Abdel-Nasser M, Mahmoud K, Omer OA, Lehtonen M, Puig D (2020) Link quality prediction in wireless community networks using deep recurrent neural networks. Alexandria Eng J 59(5):3531–43
2. Adamic L (2005) The Political Blogosphere and the 2004 U.S. Election: Divided They Blog. Proceedings of the 3rd International workshop on link discovery
3. Adamic LA, Adar E (2003) Friends and neighbors on the web. Soc Netw 25(3):211–30
4. Ali A, Zhu Y, Chen Q, Yu J, Cai H (2019) Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks. In: 2019 IEEE 25th International conference on parallel and distributed systems (ICPADS). IEEE, pp 125–132
5. Ali A, Zhu Y, Chen Q, Yu J, Cai H (2019) Leveraging spatio-temporal patterns for predicting citywide traffic crowd flows using deep hybrid neural networks. In: 2019 IEEE 25th International conference on parallel and distributed systems (ICPADS). IEEE, pp 125–132
6. Ali A, Zhu Y, Zakarya M (2021) Exploiting dynamic spatio-temporal correlations for citywide traffic flow prediction using attention based neural networks. Inform Sci 577:852–70
7. Ali A, Zhu Y, Zakarya M (2021) A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. Multimed Tools Appl 19:1–33
8. Ali A, Zhu Y, Zakarya M (2021) A data aggregation based approach to exploit dynamic spatio-temporal correlations for citywide crowd flows prediction in fog computing. Multimed Tools Appl, 1–33
9. Ali A, Zhu Y, Zakarya M (2022) Exploiting dynamic spatio-temporal graph convolutional neural networks for citywide traffic flows prediction. Neural Netw 145:233–47
10. Aziz F, Gul H, Muhammad I, Uddin I (2020) Link prediction using node information on local paths. Physica A Stat Mechan Appl 557:124980
11. Bae J, Kim S (2014) Identifying and ranking influential spreaders in complex networks by neighborhood coreness. Physica A Stat Mechan Appl 395:549–59
12. Batagelj V, Mrvar A (2006) Pajek datasets http://vlado.fmf.uni-lj.si/pub/networks/data/mix.USAir97.net
13. Behera DK, Das M, Swetanisha S, Nayak J, Vimal S, Naik B (2021) Follower Link Prediction Using the XGBoost Classification Model with Multiple Graph Features. Wirel Pers Commun. 1-20

14. Berahmand K, Nasiri E, Forouzandeh S, Li Y (2021) A preference random walk algorithm for link prediction through mutual influence nodes in complex networks. Journal of king saud university-computer and information sciences
15. Berahmand K, Nasiri E, Rostami M, Forouzandeh S (2021) A modified DeepWalk method for link prediction in attributed social network. Computing 103:2227–2249
16. Biswas A, Biswas B (2017) Community-based link prediction. Multimed Tools Appl 76(18):18619–39
17. Bonchi F, Castillo C, Gionis A, Jaimes A (2011) Social network analysis and mining for business applications. ACM Transactions on intelligent systems and technology (TIST) 2(3):1–37
18. Breiman L (2001) Random forests. Machine learning 45(1):5–32
19. Chebotarev P, Shamis E (2006) The matrix-forest theorem and measuring relations in small social groups. arXiv:math/0602070
20. Das D (2018) Positive and negative link prediction algorithm based on sentiment analysis in large social networks. Wireless Personal Commun 102(3):2183–98
21. Firth JA, Sheldon BC (2015) Experimental manipulation of avian social structure reveals segregation is carried over across contexts. Proc R Soc B Biol Sci 282(1802):20142350
22. Forouzandeh S, Rostami M, Berahmand K (2021) Presentation a Trust Walker for rating prediction in recommender system with Biased Random Walk: Effects of H-index centrality, similarity in items and friends. Eng Appl Artif Intell 104:104325
23. Gao M, Chen L, Li B, Liu W (2018) A link prediction algorithm based on low-rank matrix completion. Appl Intell 48(12):4531–50
24. Ghorbanzadeh H, Sheikhahmadi A, Jalili M, Sulaimany S (2021) A hybrid method of link prediction in directed graphs. Expert Systems with Applications. 165:113896
25. Gu S, Chen L, Li B, Liu W, Chen B (2019) Link prediction on signed social networks based on latent space mapping. Appl Intell 49(2):703–22
26. Haghani S, Keyvanpour MR (2019) A systemic analysis of link prediction in social network. Artif Intell Rev 52(3):1961–1995
27. Ibrahim NMA, Chen L (2015) Link prediction in dynamic social networks by integrating different types of information. Appl Intell 42(4):738–750
28. Kaya B (2020) A hotel recommendation system based on customer location: a link prediction approach. Multimedia Tools and Appl 79(3):1745–58
29. Kim J, Diesner J (2019) Formational bounds of link prediction in collaboration networks. Scientometrics 119(2):687–706
30. Kumar S, Lohia D, Pratap D, Krishna A, Panda BS (2021) MDER: Modified degree with exclusion ratio algorithm for influence maximisation in social networks. Computing, 1–24
31. Kumar S, Panda BS (2020) Identifying influential nodes in Social Networks: Neighborhood Coreness based voting approach. Physica A Stat Mechan Appl 553:124215
32. Kumar S, Panda A (2021) Identifying influential nodes in weighted complex networks using an improved WVoterank approach. Appl Intell, 1–15
33. Kumar S, Panda BS, Aggarwal D (2021) Community detection in complex networks using network embedding and gravitational search algorithm. J Intell Inform Syst 57(1):51–72
34. Kumar S, Saini M, Goel M, Panda BS (2021) Modeling information diffusion in online social networks using a modified forest-fire model. J Intell Inform Syst 56(2):355–377
35. Kumar A, Singh SS, Singh K, Biswas B (2019) Level-2 node clustering coefficient-based link prediction. Appl Intell 49(7):2762–79
36. Kumar S, Singhla L, Jindal K, Grover K, Panda BS (2021) IM-ELPR: Influence Maximization in social networks using label propagation based community structure. Appl Intell 51:7647–7665
37. Leicht EA, Holme P, Newman ME (2006) Vertex similarity in networks. Phys Rev E 73(2):026120
38. Leskovec J, Kleinberg J, Faloutsos C (2007) Graph evolution: Densification and shrink- ing diameters. ACM transactions on Knowledge Discovery from Data (TKDD), 1(1)
39. Li K, Tu L, Chai L (2020) Ensemble-model-based link prediction of complex networks. Comput Netw 166:106978
40. Liu W, Lü L (2010) Link prediction based on local random walk. EPL (Europhysics Letters) 89(5):58007
41. Martínez V, Berzal F, Cubero JC (2017) A survey of link prediction in complex networks. ACM Comput Surv (CSUR) 49(4):69
42. Mistele T, Price T, Hossenfelder S (2019) Predicting authors' citation counts and h-indices with a neural network. Scientometrics. 120(1):87–104
43. Monteserin A, Armentano MG (2019) Influence me! Predicting links to influential users. Inform Retriev J 22(1):32–54
44. Mutlu EC, Oghaz TA (2019) Review on graph feature learning and feature extraction techniques for link prediction. arXiv:1901.03425

45. Nasiri E, Berahmand K, Li Y (2021) A new link prediction in multiplex networks using topologically biased random walks. chaos, Solitons. Fractals 151:111230
46. Page L, Brin S, Motwani R, Winograd T (1999) The PageRank citation ranking, Bringing order to the web. Stanford InfoLab
47. Richardson M, Agrawal R, Domingos P (2003) Trust management for the semantic web. In: International semantic Web conference. Springer, Berlin, pp 351–368
48. Rozemberczki B, Davies R, Sarkar R, Sutton C (2019) Gemsec: Graph embedding with self clustering. In: Proceedings of the 2019 IEEE/ACM international conference on advances in social networks analysis and mining, pp 65–72
49. Salavati C, Abdollahpouri A, Manbari Z (2019) Ranking nodes in complex networks based on local structure and improving closeness centrality. Neurocomputing 336:36–45
50. Salton G, McGill MJ (1983) Introduction to modern information retrieval. McGraw-Hill, Auckland
51. Stam CJ, Reijneveld JC (2007) Graph theoretical analysis of complex networks in the brain. Nonlinear Biomedical Physics 1(1):1–9
52. Sun Z, Han L, Huang W, Wang X, Zeng X, Wang M, Yan H (2015) Recommender systems based on social networks. J Syst Softw 99:109–19
53. Tripathi SP, Yadav RK, Rai AK, Tewari RR (2019) Hybrid approach for predicting and recommending links in social networks. In: Computational Intelligence: theories, Applications and Future Directions-Volume II. Springer, Singapore, pp 107–119
54. Wahid-Ul-Ashraf A, Budka M, Musial K (2019) How to predict social relationships—Physics-inspired approach to link prediction. Physica A Stat Mechan Appl 523:1110–29
55. Wang Z, Liang J, Li R (2018) A fusion probability matrix factorization framework for link prediction. Knowl-Based Syst 159:72–85
56. Wang G, Wang Y, Li J, Liu K (2021) A Multidimensional Network Link Prediction Algorithm and Its Application for Predicting Social Relationships. Journal of Computational Science. 101358
57. Wang W, Wu L, Huang Y, Wang H, Zhu R (2019) Link prediction based on deep convolutional neural network. Information 10(5):172
58. Wang W, Wu L, Huang Y, Wang H, Zhu R (2019) Link prediction based on deep convolutional neural network. Information 10(5):172
59. Wen T, Deng Y (2020) Identification of influencers in complex networks by local information dimensionality. Inform Sci 512:549–62
60. White JG, Southgate E, Thomson JN, Brenner S (1986) The structure of the nervous system of the nematode Caenorhabditis elegans. Philos Trans R Soc Lond B Biol Sci 314(1165):1–340
61. Wu J, Shen J, Zhou B, Zhang X, Huang B (2019) General link prediction with influential node identification. Physica A Stat Mechan Appl 523:996–1007
62. Wu X, Wu J, Li Y, Zhang Q (2020) Link prediction of time-evolving network based on node ranking. Knowl-Based Syst 195:105740
63. Yadav RK, Rai AK (2020) Incorporating communities' structures in predictions of missing links. J Intell Inform Syst 55:183–205
64. Yao L, Wang L, Pan L, Yao K (2016) Link prediction based on common-neighbors for dynamic social network. Procedia Computer Science 83:82–9
65. Zeng A, Zhang CJ (2013) Ranking spreaders by decomposing complex networks. Phys Lett A 377(14):1031–1035
66. Zhang L, Zhao M, Zhao D (2020) Bipartite graph link prediction method with homogeneous nodes similarity for music recommendation. Multimed Tools Appl 79(19):13197–215
67. Zhou W, Gu J, Jia Y (2018) H-Index-based link prediction methods in citation network. Scientometrics 117(1):381–90

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

Affiliations

Sameer Anand[1] · Rahul[2] · Abhishek Mallik[2] · Sanjay Kumar[2]

Sameer Anand
sameeranand@sscbsdu.ac.in

Rahul
rahulv0530@gmail.com

Abhishek Mallik
abhishekmallik265@gmail.com

[1] Shaheed Sukhdev College of Business Studies (SSCBS), Delhi University, Dr. K.N. Katju Marg, New Delhi, 110089, India

[2] Department of Computer Science and Engineering, Delhi Technological University, Main Bawana Road, New Delhi, 110042, India

Psychol Stud
https://doi.org/10.1007/s12646-022-00654-x

RESEARCH IN PROGRESS

Interrelation Between Gratitude and Vitality Among Students: Exploring Mediating Role of Resilience

Naval Garg[1] · **Sarika Kumari**[2] · **B. K. Punia**[3]

Received: 30 October 2020 / Accepted: 22 February 2022
© The Author(s) under exclusive licence to National Academy of Psychology (NAOP) India 2022

Abstract The present study aspired to explore the interrelationship between gratitude and vitality among students. It also intended to investigate the mediating effect of resilience across the relationship between gratitude and vitality. Interrelation was studied using correlation and multiple regression. Mediation analysis was done with the help of hierarchical regression, Sobel test, and bootstrapping estimates. Results revealed a positive linkage between gratitude and vitality. Also, resilience was reported to mediate (partially) the relationship between gratitude and vitality among students.

Keywords Vitality · Gratitude · Resilience · Hierarchical regression · Sobel test · Bootstrapping

Abbreviations

LOSD	Lack of sense of deprivation
AO	Appreciation for others
SA	Simple appreciation
V	Vitality
R	Resilience

✉ Naval Garg
Naval.garg@dtu.ac.in

Sarika Kumari
sarika.kumari12343@gmail.com

B. K. Punia
bkpunia@gmail.com

[1] USME, Delhi Technological University, Delhi, India

[2] Department of Telecommunications, Delhi, India

[3] Haryana School of Business, GJUS&T, Hisar, India

Introduction

Swami Vivekananda said, 'Give me a hundred energetic youth, and I shall transform India.' Clearly, the great man was emphasizing the importance of energy or vitality for youth. Youth is the most productive phase of human life, it is the time when one builds one's career and starts contributing to the progress of the country. Youth act as an indispensable foundation stone for society. Their active and wholehearted participation is required to solve the challenges faced by any country. They have given dreams and hopes of a progressive, prosperous, and peaceful future since times immemorial. But, today's youth, especially students, are facing issues of lack of vitality and resilience that restricts them from realizing their true potential (Blackwell et al., 2020; Garg & Sarkar, 2020).

Vitality is the feeling of being alive, vigorous, and energetic (Blackwell et al., 2020). It indicates a state of abundance of physical and mental energy and stamina (Ryan & Frederick, 1997). High vitality quotient allows students to undertake every assignment with full engagement, enthusiasm, and spirit (Tummers et al., 2018). It is a significant predictor of overall motivation and well-being. This self-originating internal energy acts as a shock absorber for failures of life. An energetic person takes a lesson from failure and starts his journey afresh. Higher vitality increases the capability to rebound from any adverse life situation. Miksza et al. (2019) found that vitality positively influences students' ability to adapt to challenges and the quality of their associations with their peers. Although a sufficient amount of literature is available that concludes the benefits of high vitality quotient, researchers argue a lack of exploration of antecedents of vitality (Tummers et al., 2018). Accordingly, based on the broaden-and-build theory of positive emotions, the present

Published online: 09 May 2022

study makes a novel attempt by examining the linkage between gratitude and vitality.

Gratitude is a state of thankfulness towards others for their valuable contribution to one's life. It is an attribution-dependent state (Jans-Beken et al., 2019). It depends upon the beneficiary's ability to empathically realize the positivities associated with benefits received from some external source. The external source could be a person, animal, or nature (Fredrickson, 2001). It is proposed that gratitude positively influences vitality. When individuals realize and acknowledge the blessing in their life, they are more likely to feel energetic (Garg & Sarkar, 2020). The current study also examined the mediating effect of resilience amid the relationship of gratitude and vitality. Resilience is the ability to withstand failure, adversity, and misfortune. It is also referred to as one's capability to bounce back from a challenging and complicated situation. Previous researchers, including Garg and Sarkar (2020), and Simon and Gibson (2019) have reported a positive association between resilience and vitality.

Literature Review

Gratitude

Gratitude is defined as 'the willingness to recognize the unearned increments of value in one's experience' (Bertocci & Millard, 1963, p. 389). Also, gratitude is 'an estimate of gain coupled with the judgment that someone else is responsible for that gain' (Solomon, 1977, p. 316). These gains may enrich an individual with a materialistic gift such as money, workforce help, etc. Non-materialistic gifts like happiness, satisfaction, etc., also lead to a feeling of gratitude (Garg & Gera, 2019). Researchers have highlighted two crucial pre-conditions for realization of gratitude—a) some good has been done; b) this goodness is attributed to some external agency. Previous researchers have analyzed gratitude at three levels—*affective trait, mood, and emotional state*. Gratitude as an affective trait is a 'general tendency to recognize and respond with grateful emotions to the roles of other people's benevolence in the positive experiences and outcomes that one obtains' (McCullough et al., 2002, p. 112). The emotion of gratitude is expressed in the form of a feeling of thankfulness. As a mood, gratitude uplifts one's spirit for a brief period (McCullough et al., 2004).

Since both the emotion and the mood attributes of gratitude are transient, the present paper explored the trait attribute of gratitude. Watkins (2016) identified three major constructs of trait gratitude. Firstly, a grateful person develops a 'lack of sense of deprivation (LOSD) or a sense of abundance.' Such people appreciate an abundance of life and focus on the beauty and comprehensiveness of nature. This sense of abundance allows one to cherish the present moment (Garg, 2017). It bestows contentment and satisfaction. Secondly, gratitude confers 'a strong sense of appreciation for the simple pleasures of life' (SA). A grateful man starts appreciating every minute and insignificant things of day-to-day life such as awesome weather, rain, all of a sudden reunion with an old friend, early sunrise, a blossoming flower, etc. And lastly, thankfulness develops a wonderful trait of 'social appreciation or appreciation for others' (AO; Watkins, 2004). The trait of social appreciation allows the realization of others' contributions to one's life. On receiving benefits from benefactors, beneficiary tends to reciprocate the same either overtly or covertly.

Vitality

Vitality originated from the word 'vita' or 'life.' It is 'the experience of having energy available to oneself' (Ryan & Frederick, 1997, p. 2). Different researchers refer to vitality as 'an animating life force,' 'principle of life,' 'general energy for life,' 'spirit and spontaneity,' and 'vigor and liveliness of life (Lavrusheva, 2020; Garg & Sarkar, 2020; Gilstrap, et al. 2011).' Ryan and Frederick (1997) argue that vitality is an internal source of energy and is originated within oneself. It has both bodily and psychological implications, which call for a healthy (disease-free), vivacious, lively, and energetic body and mind (Stern, 2010). One of the greatest strengths of vitality is the contagious nature of energy. As energy moves from energy-rich sources to energy-less sources, an individual full of energy is poised to energize another person coming in their contact (Peterson & Seligman, 2004).

Resilience

According to the American Psychological Association, resilience is 'the process of adapting well in the face of adversity, trauma, tragedy, threats, or even significant sources of stress—such as family and relationship problems, serious health problems, or workplace and financial stressors. It means bouncing back from painful experiences (Bonanno, 2004). It is a process, capability, attitude, or outcome of successful recovery from an undesired event (Masten et al., 1990). As a process, resilience focuses on the pathways through which one bounces back to normality after a painful incident (Luthar et al., 2000). As one's capability emphasized the ability to adapt to challenging situations (Masten et al., 1990), Richardson (2002) elaborated that favorable results attained after overcoming a stressful event could be understood as aspects of resilience.

Gratitude, Resilience, and Vitality

The association between gratitude, resilience, and vitality can be understood in the background of Fredrickson's (1998) 'broaden-and-build theory of positive emotion.' The theory states that positive emotions such as gratitude 'broaden people's momentary thought-action repertoires and build their enduring social resources' (Fredrickson, 2004, p. 147). This broadening of thought-action tendencies elicits a well-thought response from the beneficiary. Also, positive emotions tend to enlarge one's scope of attention and a range of cognition (Watkins, 2016). These extensions expand one's resilience capabilities. Introspection and cognitive stretching enhance mental strength and psychological toughness, which empowers an individual. This person can efficiently deal with traumatic situations, and their psychological toughness helps in early rebounding. In this way, gratitude may be proposed as an antecedent of resilience.

Now, according to the broaden-and-build theory, gratitude can broaden one's resilience, which may build a resource of vitality for a person. Successful recovery from an undesired event boosts confidence and morale,nd it is a well-known fact that confidence energizes. Fredrickson argued that gratitude builds personal, psychological, and social resources for an individual. Here, it is proposed that gratitude leads to vitality, which is an essential psychological resource, especially for students. A person high on the vitality parameter is full of enthusiasm, vigor, and energy. Thus, it is proposed that gratitude broadens one's resilience (ability to bounce back), which leads to enhanced vitality among students. It theoretically explains the mediating effect of resilience amid the relationship of gratitude and vitality. A confident man is full of enthusiasm, vigor, and energy. It theoretically explains the mediating effect of resilience amid the relationship of gratitude and vitality.

Previous researchers reported a positive association between gratitude and several positive emotions like happiness, compassion, love, and vitality (Emmons & Shelton, 2002). McCullough et al. (2002) also reported a healthy relationship between gratitude and vitality in young adults. Various other scholars such as Christopher and Schlenker (2004), Kasser and Kasser (2001), and Garg et al. (2020) also supported the linkage of gratitude and vitality. The connection between resilience and vitality is based on the 'Job-Demands-Resources (JDR) model' (Bakker & Demerouti, 2007; Demerouti et al., 2001). This theory provides a framework that establishes resilience as an antecedent of vitality. Previous studies such as Avey et al. (2009) and Avey et al. (2011) also confirmed the claim of the Job-Demands-Resources model. Understanding the association between these two constructs is crucial because they are critical for developing critical thinking, leadership skills, and holistic development of personality (Antunes & Franco, 2016).

Although few studies reported linkage between gratitude and resilience and resilience and vitality, research in Indian settings is scarce. Most of the previously reported studies were carried out in non-eastern conditions. Researchers recommend explorations of these affiliations in eastern settings for cross-validation of results. Based on these premises, the following hypothesized model is proposed (Fig. 1).

Methodological Framework

Sample

The study's sample size is 350 students enrolled in various educational institutes in northern India. This sample comprised 224 (64%) male students and 126 (36%) female students. The age of these students varied from 18 to 28 years, with a mean value of 23.1 years. Also, 170 students were enrolled in undergraduate courses, 135 were studying in postgraduate courses, and the remaining 45 were pursuing MPhil or Ph.D. courses. The respondents were given a structured questionnaire through a google form. The purpose of the study was depicted on the first page of the google form, and informed consent of respondents was sought through a tick box. Respondents were promised that collected information would be used for academic purposes only.

Measures

Gratitude was accessed using Watkins et al. (2003)'s 16-items Gratitude Resentment and Appreciation Test short form (GRAT). This scale is a 9-point rating scale ranging from 1 (strongly disagree) to 9 (strongly agree). It measured three constructs of gratitude–LOSD, AO, and SA. The scale comprised statements such as 'I don't think that I've gotten all the good things that I deserve in life,' 'Every Fall I really enjoy watching the leaves change colors,' and 'Although I'm basically in control of my life, I can't help but think about all those who have supported me and helped me along the way.' The scale has been chosen for two specific reasons. Firstly, it is a multidimensional scale that explores three constructs of gratitude. Secondly, the scale has been validated in the Indian context (Garg & Mehak, 2021).

Vitality was measured with the help of Ryan and Frederick's (1997) 7-item Subjective Vitality Scale. It measured one's vitality on a 7-point rating scale that ranges from 1 (not at all true) to 7 (very true). It included

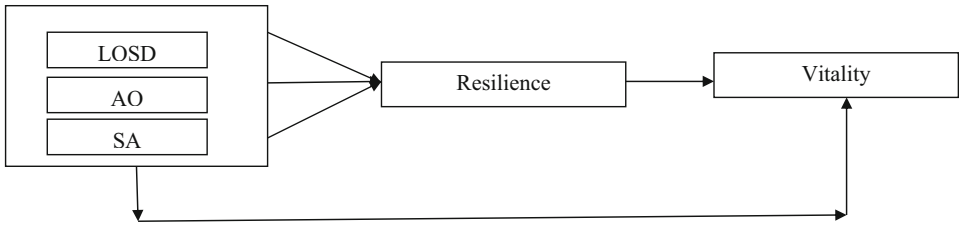


Fig. 1 Proposed Mediation Model

statements such as 'I look forward to each new day,' 'Sometimes I am so alive I just want to burst,' and 'I nearly always feel awake and alert.'

Resilience was captured using Smith et al. (2008)'s 6-item Brief Resilience Scale. This scale accessed one's inherent capability to bounce back from a traumatic incident. It is a 5-point rating scale ranging from 1 (strongly disagree) to 5 (strongly agree). These items included, 'It does not take me long to recover from a stressful event,' 'I tend to take a long time to get over setbacks in my life,' and 'I tend to bounce back quickly after hard times.'

Reliability, Validity, and Multicollinearity

Reliability was statistically calculated with the help of Cronbach's alpha (CA), whose threshold value is 0.70 (George & Mallery, 2002). Here, Cronbach's alpha values for LOSD, AO, and SA and gratitude scale were reported as 0.823, 0.811, 0.788, and 0.812, respectively. These values confirm the reliability of the gratitude data. The vitality scale showed very high reliability with a CA value of 0.912. The resilience scale was also reliable for Indian students with a CA of 0.876. Further, the Average Variance Extracted (AVE) was calculated to establish convergent validity. The AVE values for LOSD, AO, SA, and gratitude were 0.712, 0.698, 0.655, and 0.693. Fornell and Larcker (1981) argued that the AVE value greater than 0.50 assured the validity of the data. Acceptable AVE values were also reported for vitality (0.634) and resilience (0.645). Variance Inflation Factor (VIF) values less than five may indicate the absence of statistical issues caused by multicollinearity (please refer to Table 1).

Statistical Method

The interrelationship between gratitude and vitality was explored with the help of correlation and multiple regressions. Also, the mediating effect of resilience was analyzed using hierarchical multiple regression, Sobel test, and bootstrapping estimates. Hierarchical multiple regression was used for mediation analysis using Baron and Kenny's (1986) four-step process. The first step included a significant association between independent variables (LOSD, AO, and SA) and the dependent variable (vitality). The second stage required a considerable relationship between independent variables and the mediator (resilience). The third step included a significant relationship between the mediator (resilience) and the dependent variable (vitality). In the last step, the mediator was controlled, and the association between predictor and criterion variables was calculated. Here, the insignificant relationship confirmed full mediation, and a significant reduction in the regression coefficient indicated partial mediation.

Result and Analysis

Table 1 reports mean, standard deviation, CA, AVE, and VIF estimates. As discussed, these values accessed the psychometric properties of various scales used in the study. Table 2 reports the correlation between three constructs of gratitude and vitality and resilience. It concluded a statistically significant correlation at the 0.01 level. The correlation only indicates the degree and direction of the association between variables. Regression allows the

Table 1 Descriptive Statistics

Variable	Mean	S.D	A	AVE	VIF
LOSD	6.21	1.432	0.823	0.712	3.451
AO	5.98	0.783	0.811	0.698	3.237
SA	6.16	0.973	0.788	0.655	4.472
Gratitude	6.11	0.854	0.812	0.693	–
Vitality	4.88	1.132	0.912	0.634	–
Resilience	3.17	0.733	0.876	0.631	–

Source: Primary Data, *S.D* standard deviation, α Cronbach's alpha, *AVE* average variance extracted, *VIF* variance inflation index

Table 2 Correlation matrix

Variable	1	2	3	4	5
LOSD	1				
AO	0.362*	1			
SA	0.352*	0.285*	1		
Vitality	0.587*	0.649*	0.621*	1	
Resilience	0.489*	0.473*	0.563*	0.580*	1

Source: Primary Data, *Correlation is significant at the 0.01 level

prediction of the dependent variable based on independent variables. Accordingly, three constructs of gratitude were regressed over vitality.

Table 3 elaborates result of the regression model ($V = \beta_0 + \beta_1$ LOSD $+ \beta_2$ AO $+ \beta_3$ SA) that considered vitality as a dependent variable and LOSD, AO, and SA as independent variables. Result reported that these constructs of gratitude positively explained variations in the vitality of students ($\beta_{LOSD} = 0.546$, p value < 0.001; $\beta_{AO} = 0.634$, p value < 0.001; $\beta_{SA} = 0.631$, p value < 0.001). Table 4 reports that these three constructs of gratitude explained as high as 54% variation in vitality. The following regression model was extracted:-

$$V = 3.625 + 0.546\ \text{LOSD} + 0.634\ \text{AO} + 0.631\ \text{SA}$$

Table 5 evaluates the mediating effect of resilience amid the association between gratitude and student vitality using Baron and Kenny's (1986) technique. Their mediation analysis technique is comprised of four steps. A statistically significant relationship between three independent and dependent variables, between predictor variables and mediator, and between the mediator and criterion variable at a 0.01 level of significance confirmed observance of the first three conditions of Baron and Kenny (1986). Now, on the controlling mediator, the regression coefficient decreased for all three mediation models, which concluded the partial mediation effect of resilience. The Sobel test concluded that the partial mediation effect of resilience was

Table 4 Model summary

Model	R	R Square	Adjusted R Square	Std. Error of Estimate	F	Sig
1	0.735	0.540	0.523	0.764	22.145	0.000*

Source: Primary Data, *sig. at 0.01

statistically significant. The bootstrapping results showed that zero does not lie between the LLCI and ULCI (confidence interval). It confirmed the significance of the mediation effect of resilience in all three proposed hypothesized relationships (please refer to Table 6).

Discussion and Conclusion

This study evaluated the connection between three constructs of gratitude and vitality among Indian students. It also explored the mediating effect of the virtue of resilience amid the relationship between these variables. Results suggested that all three constructs of gratitude may elicit vitality among students. It could be inferred from these findings that the LOSD develops and reaffirms positivity among students. Such students tend to focus less on life failures and are less jealous of others. This acute focus on positivity saves a lot of energy for the students, and thus, it vitalizes them. Appreciation for others may include students' gratefulness for society, institute, teachers, friends, and family for their invaluable contribution in their lives. It prompts a sense of belongingness and social affiliation with these well-wishers of life. A healthy social group is vital for proper guidance and direction, which ultimately channelizes one's energy in a positive direction. Lastly, a simple appreciation for the minute pleasures of life reduces negative emotions such as depression, and anxiety about present assignments, supervisors, and institutes. Results also empirically concluded that resilience partially mediated the connection between three constructs of gratitude and vitality. It means that a lack of sense of deprivation results in a higher level of

Table 3 Multiple regression analysis (Dependent Variable—Vitality)

Model	Unstandardized Coefficients		Standardized Coefficients	t-value	Sig
	B	Std. Error	B		
Constant	3.625	0.893		5.673	.000*
LOSD	0.512	0.262	0.546	6.356	.000*
AO	0.628	0.089	0.634	8.940	.000*
SA	0.590	0.830	0.631	6.647	.000*

Source: Primary Data, *sig. at 0.01

Table 5 Mediation analysis (Results of Hierarchical Regression)

Hypothesized relationships	IV to DV	IV to Med	M to DV	IV to DV (with a mediator)
LOSD → R → V	0.532*	0.470*	0.556*	0.384*
AO → R → V	0.602*	0.434*	0.556*	0.493*
SA → R → V	0.624*	0.535*	0.556*	0.455*

Source: Primary Data, *sig at .01, *IV* independent variable, *DV* dependent variable, *M* mediator

Table 6 Mediation analysis (Results of Sobel Test and Bootstrapping)

Hypothesized relationships	Sobel test		Bootstrapping estimates			
	z value	Sig	Effect	SE	LLCI	ULCI
LOSD → R → V	5.479	.000*	0.1526	0.036	0.0374	0.1384
AO → R → V	6.645	.000*	0.1880	0.066	0.3632	0.5399
SA → R → V	6.748	.000*	0.1182	0.025	0.2739	0.4838

Source: Primary Data, *p sig at .01

resilience, and this ultimately leads to vitality among students. It is a well-known fact that gratitude enriches one with a positive outlook for life (McCullough et al., 2002). And this positive outlook is a necessary component for developing an ability to bounce back from a traumatized situation (resilience). This will to win gives immense energy to overcome all hurdles of life.

The study adds to the existing knowledge of gratitude, resilience, and vitality. First, this study concludes a healthy connection of gratitude with two indispensable virtues for students, i.e., resilience and vitality. Second, this study tried to answer the 'black-box' of pathways through which gratitude leads to vitality. Results established resilience as a partial mediator, which transmits the effect of gratitude to vitality. Thirdly, although the relationship between gratitude and resilience and gratitude and vitality was previously explored, most previous studies were conducted in western contexts. The findings of the present study validate the results in the eastern context. Various previous researchers like Garg and Sarkar (2020), Kasser and Kasser (2001), Wilson (2016), and Kumar and Dixit (2014) also recommended the same results. Kumar and Dixit (2014) reported a significant correlation of 0.345 between gratitude and resilience. Similarly, Wilson (2016) also concluded that practicing gratitude helps students in their ability to focus and remain resilient in challenging times.

The study offers several suggestions for academicians and educational policymakers. First, teachers and academic leaders need to develop a sense of gratefulness. Since emotions are contagious, teachers' emotions are doubly contagious for students. Thus, the thankful expression of teachers may encourage students to develop a sense of abundance and simple appreciation, which may ultimately lead to highly resilient and energized students. Second, results suggest that educational institutes may adapt several gratitude-based interventions to cultivate gratitude among students. These interventions include 'gratitude journals,' 'gratitude letters,' 'gratitude collages,' 'sharing positive events,' and 'acts of kindness.' These workable inferences may easily be practiced with the help of seminars, courses, conferences, and workshops. Also, the government of India is expressing a keen interest in ensuring the holistic development of students. Several key initiatives, including happiness courses, Yoga and mediation sessions, personality development courses, the choice-based credit system, etc., have been started. It is suggested that the government encourage theoretical and practical learning and exploration of one of the most desirable human traits like gratitude. Adequate funding and encouragement through liberal research funding to researchers and training to instructors will also be beneficial.

The current study offered a practical model of gratitude, resilience, and vitality. However, the study has certain limitations too. First, the current study is based on cross-sectional data, wherein independent, mediator, and dependent variables were measured at a single point of time. Future researchers may imply longitudinal or double source studies to negate common method bias. Second, the sample for the present study was only 350 students. Future researchers may rely on a larger sample for greater reliability and generalizability of results. Third, the single-dimensional scale of vitality and resilience was used in this study. Multifaceted

measurement instruments of these variables will provide deeper insight for future researchers in the field.

Author Contributions All authors contributed to the conceptualization and designing of the study. Dr. Garg and Prof. Punia mainly designed the study, collected and analyzed the data. Mrs. Kumari provided literature and discussion.

Funding None.

Data Availability Respondents were assured about confidentiality of the data. So, author possess the data but will not be shared.

Declarations

Conflict of interest All the authors that they have no conflict of interest.

Consent to Participate Yes, all respondents were duly informed of academic usage of the collected data.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki Declaration and its later amendments or comparable ethical standards.

References

Antunes, A., & Franco, M. (2016). How people in organizations make sense of responsible leadership practices. *Leadership and Organization Development Journal, 37*(1), 126–152. https://doi.org/10.1108/lodj-04-2014-0084

Avey, J. B., Luthans, F., & Jensen, S. M. (2009). Psychological capital: A positive resource for combating employee stress and turnover. *Human Resource Management, 48*(5), 677–693. https://doi.org/10.1002/hrm.20294

Avey, J. B., Reichard, R. J., Luthans, F., & Mhatre, K. H. (2011). Meta-analysis of the impact of positive psychological capital on employee attitudes, behaviors, and performance. *Human Resource Development Quarterly, 22*(2), 127–152. https://doi.org/10.1002/hrdq.20070

Bakker, A. B., & Demerouti, E. (2007). The job demands-resources model: State of the art. *Journal of Managerial Psychology, 22*(3), 309–328. https://doi.org/10.1108/02683940710733115

Baron, R. M., & Kenny, D. A. (1986). The moderator-mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personality and Social Psychology, 51*(6), 1173–1182. https://doi.org/10.1037/0022-3514.51.6.1173

Bertocci, P. A., & Millard, R. M. (1963). *Personality and the good: Psychological and ethical perspectives*. New York, NY: David McKay.

Blackwell, J., Miksza, P., Evans, P., & McPherson, G. E. (2020). Student vitality, teacher engagement, and rapport in studio music instruction. *Frontiers in Psychology*. https://doi.org/10.3389/fpsyg.2020.01007

Bonanno, G. A. (2004). Loss, trauma, and human resilience: Have we underestimated the human capacity to thrive after extremely aversive events? *American Psychologist*, *59*(1), 20–28. https://doi.org/10.1037/0003-066X.59.1.20.

Christopher, A. N., & Schlenker, B. R. (2004). Materialism and affect: The role of self-presentational concerns. *Journal of Social and Clinical Psychology, 23*(2), 260–272. https://doi.org/10.1521/jscp.23.2.260.31022

Demerouti, E., Bakker, A. B., Nachreiner, F., & Schaufeli, W. B. (2001). The job demands-resources model of burnout. *Journal of Applied Psychology, 86*(3), 499–512. https://doi.org/10.1037/0021-9010.86.3.499

Emmons, R. A., & Shelton, C. M. (2002). Gratitude and the science of positive psychology. In C. R. Snyder & S. J. Lopez (Eds.), *Handbook of positive psychology* (pp. 459–471). Oxford University Press.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39. https://doi.org/10.2307/3151312

Fredrickson, B. L. (1998). What good are positive emotions? *Review of General Psychology, 2*(3), 300–319. https://doi.org/10.1037/1089-2680.2.3.300

Fredrickson, B. L. (2001). The role of positive emotions in positive psychology: The broaden-and-build theory of positive emotions. *American Psychologist, 56*(3), 218–226. https://doi.org/10.1037/0003-066x.56.3.218

Fredrickson, B. L. (2004). The broaden–and–build theory of positive emotions. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1449), 1367–1377. https://doi.org/10.1098/rstb.2004.1512

Garg, N. (2017). Workplace spirituality and organizational performance in indian context: Mediating effect of organizational commitment, work motivation, and employee engagement. *South Asian Journal of Human Resources Management, 4*(2), 191–211. https://doi.org/10.1177/2322093717736134

Garg, N., & Gera, S. (2019). Gratitude and leadership in higher education institutions. *Journal of Applied Research in Higher Education, 12*(5), 915–926. https://doi.org/10.1108/jarhe-09-2019-0241

Garg, N., Jain, A., & Punia, B. (2020). Gratitude, social intelligence, and leadership among university teachers: Mediation and moderation analysis. *International Journal of Organizational Analysis, 29*(2), 368–388. https://doi.org/10.1108/ijoa-01-2020-2016

Garg, N., & Mehak. (2021). Gratitude resentment and appreciation scale (GRAT-16): analyzing psychometrics properties in the Indian context. *Current Psychology*. https://doi.org/10.1007/s12144-021-02169-6

Garg, N., & Sarkar, A. (2020). Vitality among university students: Exploring the role of gratitude and resilience. *Journal of Organizational Effectiveness: People and Performance, 7*(3), 321–337. https://doi.org/10.1108/joepp-03-2020-0041

George, D., Mallery, P. (2002). *SPSS for Windows Step by Step: A Simple Guide and Reference, 11.0 Update (4th Edition)* (4th ed.). Allyn & Bacon.

Gilstrap, J. B., Harvey, J., Novicevic, M. M., & Buckley, M. R. (2011). Research vitality as sustained excellence: what keeps the plates spinning? *Career Development International*, *16*(6), 616–644. https://doi.org/10.1108/13620431111117835.

Jans-Beken, L., Jacobs, N., Janssens, M., Peeters, S., Reijnders, J., Lechner, L., & Lataster, J. (2019). Gratitude and health: An updated review. *The Journal of Positive Psychology, 15*(6), 743–782. https://doi.org/10.1080/17439760.2019.1651888

Kasser, T., & Kasser, V. G. (2001). The dreams of people high and low in materialism. *Journal of Economic Psychology, 22*(6), 693–719. https://doi.org/10.1016/s0167-4870(01)00055-1

Kumar, A., & Dixit, V. (2014). Forgiveness, gratitude, and resilience among Indian youth. *Indian Journal Of Health And Wellbeing, 5*(12), 1414–1419.

Lavrusheva, O. (2020). The concept of vitality. Review of the vitality-related research domain. *New Ideas in Psychology*, *56*, 100752. https://doi.org/10.1016/j.newideapsych.2019.100752.

Luthar, S. S., Cicchetti, D., & Becker, B. (2000). The construct of resilience: A critical evaluation and guidelines for future work. *Child Development, 71*(3), 543–562. https://doi.org/10.1111/1467-8624.00164

Masten, A. S., Best, K. M., & Garmezy, N. (1990). Resilience and development: Contributions from the study of children who overcome adversity. *Development and Psychopathology, 2*(4), 425–444. https://doi.org/10.1017/s0954579400005812

McCullough, M. E., Emmons, R. A., & Tsang, J. A. (2002). The grateful disposition: A conceptual and empirical topography. *Journal of Personality and Social Psychology, 82*(1), 112–127. https://doi.org/10.1037/0022-3514.82.1.112

McCullough, M. E., Tsang, J. A., & Emmons, R. A. (2004). Gratitude in intermediate affective terrain: Links of grateful moods to individual differences and daily emotional experience. *Journal of Personality and Social Psychology, 86*(2), 295–309. https://doi.org/10.1037/0022-3514.86.2.295

Miksza, P., Evans, P., & McPherson, G. E. (2019). Wellness among university-level music students: A study of the predictors of subjective vitality. *Musicae Scientiae, 25*(2), 143–160. https://doi.org/10.1177/1029864919860554

Peterson, C., & Seligman, M. (2004). *Character Strengths and Virtues: A Handbook and Classification* (1st ed.). American Psychological Association / Oxford University Press.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*(1), 539–569. https://doi.org/10.1146/annurev-psych-120710-100452

Richardson, G. E. (2002). The metatheory of resilience and resiliency. *Journal of Clinical Psychology, 58*(3), 307–321. https://doi.org/10.1002/jclp.10020

Ryan, R. M., & Frederick, C. (1997). On energy, personality, and health: Subjective vitality as a dynamic reflection of well-being. *Journal of Personality, 65*(3), 529–565. https://doi.org/10.1111/j.1467-6494.1997.tb00326.x

Simon, S., & Gibson, M. T. (2019). Principal resilience and vitality in extremis. *International Journal of Educational Management, 33*(4), 709–720. https://doi.org/10.1108/ijem-05-2018-0163

Smith, B. W., Dalen, J., Wiggins, K., Tooley, E., Christopher, P., & Bernard, J. (2008). The brief resilience scale: Assessing the ability to bounce back. *International Journal of Behavioral Medicine, 15*(3), 194–200. https://doi.org/10.1080/10705500802222972

Solomon, R. C. (1977). *The passions*. Anchor Books.

Stern, D. (2010). The issue of vitality. *Nordic Journal of Music Therapy, 19*(2), 88–102. https://doi.org/10.1080/08098131.2010.497634

Tummers, L., Steijn, B., Nevicka, B., & Heerema, M. (2018). The effects of leadership and job autonomy on vitality: Survey and experimental evidence. *Review of Public Personnel Administration, 38*(3), 355–377. https://doi.org/10.1177/0734371x16671980

Watkins, P. C. (2016). *Gratitude and the Good Life: Toward a Psychology of Appreciation* (Softcover reprint of the original 1st ed. 2014 ed.). Springer.

Watkins, P. C. (2004). Gratitude and Subjective Well-Being. In R. A. Emmons & M. E. McCullough (Eds.), *The psychology of gratitude* (pp. 167–192). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195150100.003.0009

Watkins, P. C., Woodward, K., Stone, T., & Kolts, R. L. (2003). Gratitude and happiness: Development of a measure of gratitude, and relationships with subjective well-being. *Social Behavior and Personality: An International Journal, 31*(5), 431–451. https://doi.org/10.2224/sbp.2003.31.5.431

Wilson, J. T. (2016). Brightening the mind: The impact of practicing gratitude on focus and resilience in learning. *Journal of the Scholarship of Teaching and Learning, 16*(4), 1–13. https://doi.org/10.14434/josotl.v16i4.19998

Yay, M. (2017). The mediation analysis with the sobel test and the percentile bootstrap. *International Journal of Management and Applied Science (IJMAS), 3*(2), 123–125.

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

Check for updates

Investigating Exhaust Emissions from In-Use Passenger Cars: Exploratory Analysis and Policy Outlook

Abhinav Pandey[1]; Rajeev Kumar Mishra[2]; and Govind Pandey[3]

Abstract: Most developing countries still have various makes of petrol-driven cars dominating the overall passenger vehicle fleet. In such countries, the emission certification policy for in-use vehicles remains an area of concern, making the inspection and maintenance (I/M) program less effective. A thorough investigation of the exhaust emissions from such cars is required to explore and address this concern. This paper provides insights into the effects of vehicle variables on tailpipe emission parameters from an exclusively larger and heterogeneous dataset of in-use cars (n = 1,580). Results showed that both vehicle variables such as age, mileage, emissions norm, and maintenance category, and two engine variables, i.e., aspiration type and fuel mixing conditions, had a significant and direct influence on tailpipe parameters, namely, CO, HC, CO_2, O_2, λ, and AFR (carbon monoxide, hydrocarbon, carbon dioxide, oxygen, lambda and air-fuel ratio respectively). Stronger correlations were found for the relatively larger (considering age, R^2 for $CO_{idle} = 0.88$, $HC_{idle} = 0.73$, $\lambda_{f.idle} = 0.74$, $AFR_{f.idle} = 0.73$ and considering mileage, R^2 for $CO_{idle} = 0.75$, $HC_{idle} = 0.67$, $\lambda_{f.idle} = 0.62$, $AFR_{f.idle} = 0.61$ for the whole dataset) and diverse make-wise (R^2 values fared even better, 0.87–0.93 for CO and 0.69–0.77 for HC) data collected during the study. The present research provides a first-hand and comprehensive analysis of the effects of the stringency of emission norms and maintenance category on the exhaust emissions from in-use cars. The polynomial emission equations generated by this study can reliably predict the emission levels for CO and HC based on the age and/or mileage of cars. Further, the results recommend revised policies to upgrade the existing emission certification infrastructure and phasing out policy of cars. **DOI: 10.1061/(ASCE)EE.1943-7870.0002015.** © *2022 American Society of Civil Engineers.*

Author keywords: Emission compliance; Idle testing; Maintenance category; Inspection/maintenance (I/M) program; Tailpipe emission; Vehicle variables; Developing countries.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Introduction

As most city areas are transforming into urban conglomerations, the rise in motor vehicle density is following closely with global expansion. This scenario (particularly in developing countries) is giving rise to the aggregation of a larger number of motor vehicles and their lifetime spent in the same urban boundary limits, thereby causing tremendous pressure on the well-being of the urban environment. These urban areas emit a significant proportion of air pollutants globally and can be often associated with poor air quality (Lawrence et al. 2007; Butler 2013).

Over the last 10–15 years, worldwide passenger car sales have embraced a very high growth of 17.97% to 21.16% (comparing the sales data of 2019 with 2010 and 2005, respectively). The 45.25 million passenger cars sold globally in 2005 have almost doubled to 90.42 million in only 14 years (OICA 2019). The vehicular growth rate has surpassed the urbanization rate, with more than 90% of the daily vehicular sales linked to private ownership (Perappadan 2012; Government of Delhi 2020).

In such countries, increasing ambient air pollution has caused over 4.2 million deaths annually, along with many cases of respiratory illnesses (WHO 2019). Furthermore, exposure to vehicular emissions (VE) have also been reported to prompt various human health consequences including, but not limited to, cardiovascular disease, pulmonary function decline, cancer, and mortality (Ogunseye et al. 2018; Rice et al. 2015; Wong et al. 2019).

Megacities, particularly in developing nations, have reported that over 70%–80% of air pollution is attributed to vehicular emissions caused by a large chunk of older vehicles exhibiting poor vehicle maintenance, inadequate road infrastructure, and low fuel quality (Badami 2005; Singh et al. 2007; Wang et al. 2010; Pandey et al. 2016). Among these pollutants, CO (carbon monoxide) is one of the most significant pollutants emitted by the transportation segment, contributing to about 90% of the total emissions, with hydrocarbons (HCs) following closely.

A heterogeneous fleet of cars in Delhi is subject to periodic emissions compliance testing, which allows them to remain in operation only upon passing the test. This testing is an integral part of the existing inspection/maintenance (I/M) program, which consists of only non-loaded idle or fast idle measurements of CO and HC from the vehicle's tailpipe while stationary. Earlier studies have expressed the need to understand the emission characteristics of vehicles with a larger, heterogeneous dataset (Kazopoulo et al. 2005) with more model years (Beydoun and Guldmann 2006) and within the context of more parameters concerning vehicles, age, mileage, maintenance, applicable emission norms, and various engine-related features (Pandey et al. 2016).

Analysis of test failure and detailed regression analyses of I/M data for CO and HC from various states in the United States concerning vehicle features and emissions revealed that the likely failure rate of older and poorly maintained vehicles was significantly higher (Beydoun and Guldmann 2006). A study on the Japanese

[1]Research Scholar, Dept. of Environmental Engineering, Delhi Technological Univ., Delhi 110 042, India.

[2]Assistant Professor, Dept. of Environmental Engineering, Delhi Technological Univ., Delhi 110 042, India (corresponding author). Email: rajeevkumarmishra@dtu.ac.in

[3]Professor, Dept. of Civil Engineering, Madan Mohan Malaviya Univ. of Technology, Gorakhpur, Uttar Pradesh 273 010, India.

Note. This manuscript was submitted on November 25, 2021; approved on February 22, 2022; published online on May 6, 2022. Discussion period open until October 6, 2022; separate discussions must be submitted for individual papers. This paper is part of the ***Journal of Environmental Engineering***, © ASCE, ISSN 0733-9372.

scrappage scheme found that CO_2 (carbon dioxide) emissions would only decrease if users retained their new gasoline passenger vehicles for less than 4.7 years. This decrease was predominantly attributable to the better combustion efficiency of newer vehicles (Kagawa et al. 2013).

In the case of an Italian fleet of passenger cars, annual mileage was found to drop significantly with age. Both diesel and gasoline cars drove half the annual distance when they reached an average age of approximately 8 years; hence, mileage must be considered along with vehicle age when estimating emissions from the transportation sector (Caserini et al. 2013).

An investigation of the exhaust emissions (CO and HC) of 100 gasoline cars under a basic I/M program in Lebanon was conducted. The vehicles reported higher failure rates indicating the need to develop country-specific emission standards. Proposed Lebanese standards had CO and HC ranges of $1.5\% \pm 0.5\%$ and 250 ± 36 (ppm), respectively. The smaller sample size was a major limitation to the study, implying the need of representative sampling to refine the outcomes (Kazopoulo et al. 2005).

During idle operation of vehicles, the engine does not operate at peak temperatures resulting in high brake-specific fuel consumption and incomplete fuel combustion that lead to high emission formation (both CO and HC) as well as fuel residue in the exhaust (Sanchita et al. 2014).

An investigation of tailpipe CO and HC from 300 petrol-driven passenger cars of Maruti make in India found both CO and HC to be higher in cases of fast idling conditions. In both test scenarios, the tailpipe emissions of CO and HC were positively correlated with vehicle age and mileage. However, the need to refine this correlation was suggested with a more intensive and make-wise dataset (Pandey et al. 2016).

A direct relationship between AFR and λ in the form of polynomial equations with high accuracy was determined based on field data and measurements. It can help in understanding the relationship between AFR and λ and determine the value of any one of them if the other value is known using the estimation equations developed using data analysis (Al-Arkawazi 2019).

Using the significant model coefficients and corresponding log odds values, a probability model was constructed to show the probability of compliance to the emission norms. For instance, a 15 year old, private car used in Ogun State, Nigeria, had a 79% probability of complying with Euro II (2.5%) standard. This probability can then be extrapolated for any aged vehicle in-use (Moonsammy et al. 2021).

The extent of the reduction in NO (nitrogen oxide) emissions by each successive Euro standard was demonstrated clearly. When the proportion of each category in the overall observed fleet was taken into account, the benefit of moving from earlier to later emissions standards to reduce overall NO emissions from petrol cars was found to be clear (Rhys-Tyler et al. 2011).

A comprehensive inspection and remedial maintenance program is extremely important to curb and control vehicular emission levels. Accordingly, like many other countries, most Asian countries have taken various initiatives to implement an effective inspection and maintenance program (Dandapat et al. 2020).

For the most part, I/M programs have two types of emissions testing, e.g., basic and advanced. The most common being the measurement of CO and HC values in the exhaust while the vehicle is idling. The idle test was originally developed with a view to detect a large proportion of malfunctioning or maladjusted engine. To help reduce the false failures with this test, some I/M programs require pre-conditioning at 2,500 revolutions per minute (rpm) with no load, referred to as fast idle testing (Pandey et al. 2016).

In spite of its introduction many years ago, idle testing has some advantages. The idle test is capable of monitoring gross emitters, and the basic idle testing is cost-effective and hence widely prevalent in developing countries, compared to the more sophisticated equipment used in advanced tests. Idle mode emissions for CO and HC are high compared to those of other driving modes and idling, as well as low-speed ranges, occupy a large proportion of total driving time in urban areas.

In view of the advantages offered by idle testing, the I/M program framework in many developing countries still relies on idle (low or high) testing for in-use emissions compliance by petrol-driven cars. Further, as the NCT (National Capital Territory) of Delhi is characterized by several congested traffic intersections where 15%–20% of the daily commute is wasted in idling. Therefore, both idle and fast idle tests were performed in this study, even though the correlation between these two test results and emissions measured under more realistic driving conditions is poor (Pandey et al. 2016).

As part of the Indian I/M program, all in-use vehicles are required to mandatorily obtain a Pollution Under Control (PUC) certificate, which is issued based on successful compliance to idle emission values of CO and HC for petrol vehicles using auto-exhaust gas analyzers only at government-authorized centers (CPCB 2010). This system has several flaws and needs specific changes for better outcomes (Fig. 1).

From the literature review, we found that there is a need to address the following gaps:

- Most of the studies have expressed the need to refine their findings (effect of vehicle variables on emissions) with a larger, representative, and heterogeneous dataset.
- Studies attempting to conduct thorough in-situ analysis of tailpipe emissions and vehicular traits regarding emissions

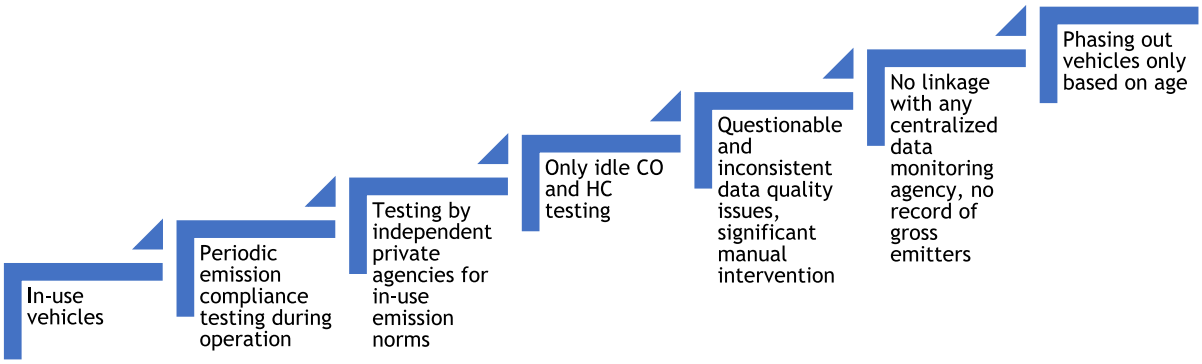


Fig. 1. Existing emissions compliance testing framework for in-use cars.

J. Environ. Eng., 2022, 148(7): 04022035

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

compliance testing (especially in developing countries) are highly required.
- Studies on the effects of the progressive stringency of emission norms on vehicular emissions, particularly in-use vehicles are scantily available and need to be conducted to bring better understanding.
- As the in-use vehicles form a substantial portion of the overall vehicular fleet, an assessment of their actual emissions, vis-à-vis vehicle variables, is required to ascertain the changes required in the existing I/M framework.

Therefore, the present study assessed the petrol-driven passenger cars appearing at PUC centers for emissions recertification for their tailpipe parameters in a collaborated testing program. Based on the data analysis, useful inferences are drawn indicating the need to strengthen the I/M program in developing countries.

Materials and Methods

Emission Testing Program

The testing of vehicles was carried out in collaboration with PUC centers located across various districts of the NCT (Fig. S1). The vehicle testing program included 1,580 petrol-driven passenger cars. On average, 4–6 vehicles were tested daily, with only one PUC center engaged in testing at one point in time. The vehicles stopped at the test platform and were checked for tailpipe emissions in idle and fast idle conditions along with vehicle- and engine-related parameters.

Vehicle Characteristics

District- and RTO (concerned Regional Transport Office)-wise distributions of all cars showed maximum numbers belonging to the South-West district and Dwarka RTO followed by the West district and Wazirpur and Rajouri Garden RTOs representing more than half of the total sample size. The hatchback body type dominated the petrol cars (69.36%) followed by sedans (26.14%) and SUVs (Sports Utility Vehicles, 4.5%). Make-wise distribution is provided in Table 1, showing MSIL (Maruti Suzuki India Limited), HMIL (Hyundai Motors India Limited), and HCIL (Honda Cars India Limited) representing about 83.61% of the total number of cars. The petrol cars' fleet had a mix of indigenous and overseas cars makers with an almost 50:50 ratio.

Table 1. Make-wise sample frequencies for petrol cars

S. no.	Make	Country	Frequency	% age
1	Audi	Germany	1	0.063
2	FAIPL (Fiat)	Italy	1	0.063
3	FIPL (Ford)	USA	37	2.342
4	GMIPL (General Motors)	USA	25	1.582
5	HCIL (Honda)	Japan	205	12.975
6	HMIL (Hyundai)	South Korea	381	24.114
7	HML (Hindustan)	India	1	0.063
8	MML (Mahindra)	India	5	0.316
9	MSIL (Maruti)	India	735	46.519
10	NMIPL (Nissan)	Japan	7	0.443
11	RIPL (Renault)	France	25	1.582
12	SAIPL (Skoda)	Czech	9	0.570
13	TKMPL (Toyota)	Japan	56	3.544
14	TML (Tata)	India	51	3.228
15	VWIPL (Volkswagen)	Germany	41	2.595
	Total		1,580	100.00

Data Collection and Instrument Characteristics

During the emissions testing program, various dependent and independent variables were recorded. The first independent variable set included vehicle-related parameters. It was comprised of make (whole data, MSIL/HMIL/HCIL), body type (hatchback/sedan/SUV), age, mileage, emission norm at manufacturing, and maintenance category (very good/good/poor/unsatisfactory), while the second set concentrated on engine-specific aspects such as capacity, maximum power, maximum torque, compression ratio, cylinder bore, piston stroke, aspiration type, no. of cylinders, no. of valves per cylinder, valve configuration, and fuel mixing condition. The dependent variable set comprised of tailpipe emission parameters, i.e., CO, HC, CO_2, O_2, λ, and AFR.

Although the idle test helps identify gross emitters, it does not simulate real-world driving conditions. In Lebanon, idle testing was

Table 2. Field instrument characteristics—4-gas analyzer for petrol-driven cars

Item	Particulars
Make/model	Ozone/Oz-Gas-05
Government's approval for application at/by PUCCs	Yes/verified
Constituents measured	CO; CO_2; HC; O_2; AFR; λ
Measurement type	Mass concentration
Measurement range	CO–0%–15%; CO_2–0.01%; HC–1 ppm; O_2–0.01%
Measurement resolution	CO–0.001%; CO_2–0%–20%; HC–0–30,000 ppm; O_2–0%–25%
Principal of operation	$CO/CO_2/HC-NDIR\ O_2-ECS$
Operating temperature (°C)	4–45
RPM measurement range	400–9,999
Unit weight (kg)	7.5 (approximately)
Unit dimensions	344 × 368 × 223 mm (L × W × H)

Note: NDIR = non-dispersive infrared; and ECS = elctro-chemical sensor/server.

Table 3. Vehicle and engine characteristics of petrol cars

Item	Particulars
Vehicles	
Total numbers	1,580
Fuel distribution type	MPFI (multi-point fuel injection)
Vehicle age range (years)	0.01–20.04
Average age (years)	4.37
Oldest/newest vehicle tested	M-800 (MSIL)/Baleno (MSIL)
Vehicle mileage range (km)	110–200,356
Average mileage (km)	45,605
Vehicle with highest/lowest mileage (km)	M-800 (MSIL)/Baleno (MSIL)
No. of vehicles with registration life of 15 years over/half-past	3/255
Kerb weight range (kg)	635–1,855
Top grosser make	MSIL (n = 735)
Top three makes	MSIL/HMIL (n = 321)/ HCIL (n = 205)
Engines	
Capacity range (cc)	624–2,694
Maximum power range (bhp)	35–187.8
Maximum torque range (Nm)	51–320
Compression ratio range	8.8:1–17.6:1
Cylinder bore range (mm)	66–95
Piston stroke range (mm)	77–103.6
No. of cylinders/no. of valves per cylinder (range)	2–4/2–4

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

performed based on how much time the vehicles spend in traffic while the engines idled (Kazopoulo et al. 2005; Frey et al. 2003). As the NCT of Delhi is characterized by several congested traffic intersections, the vehicles passed a substantial amount of time idling during their commute (15%–20%).

After selecting the testing method(s), we tested 1,595 cars distributed across model years ranging between 1996 and 2019. The overall study methodology is outlined in Fig. S2.

The study used a four (4)-gas analyzer (Oz-Gas-05, Ozone Electronics, Delhi, India) to measure mass concentrations of tailpipe (emission) parameters, i.e., CO, CO_2, O_2 (in % terms); HC (in ppm terms) λ; and AFR (both as dimension-less entities). The instrument catered to both idle and fast idling test conditions having a hand-held digital tachometer for engine RPM measurements. The particulars of the testing instrument are given in Table 2.

Data Analysis Tools

Considering the large volume of data, the number of variables collected during the emissions testing program and the aims of the research, the study used the SPSS package (IBM SPSS Statistics, version 23) for emissions data analysis. The MS-Excel program was also used intermittently for short-term data processing. Out of the 1,595 cars tested, 12 cars were found with duplicate entries (i.e., same car inadvertently appeared at two different PUC centers). Upon a detailed review, we found the measurement values for tailpipe parameters to be consistent at both PUCs for these 12 cars. Further, we found that four cars had missing values for some of the parameters, while 11 cars returned emissions values as outliers (unreliably beyond the anticipated values of the tailpipe exhaust constituents). Hence, considering this number ($4 + 11 = 15$) to be negligible looking at the total numbers tested ($15/1{,}595 = 0.94\%$ i.e., <1%), it was removed from the final analysis numbers ($n = 1{,}580$ used in the data analysis; Fig. S2).

Results and Discussion

Tailpipe concentrations of CO, HC, CO_2, O_2, and values of λ and AFR were measured in the present work. Such values are plotted with respect to vehicle-related parameters, such as vehicle age, vehicle mileage, applicable emission norm (at manufacturing),

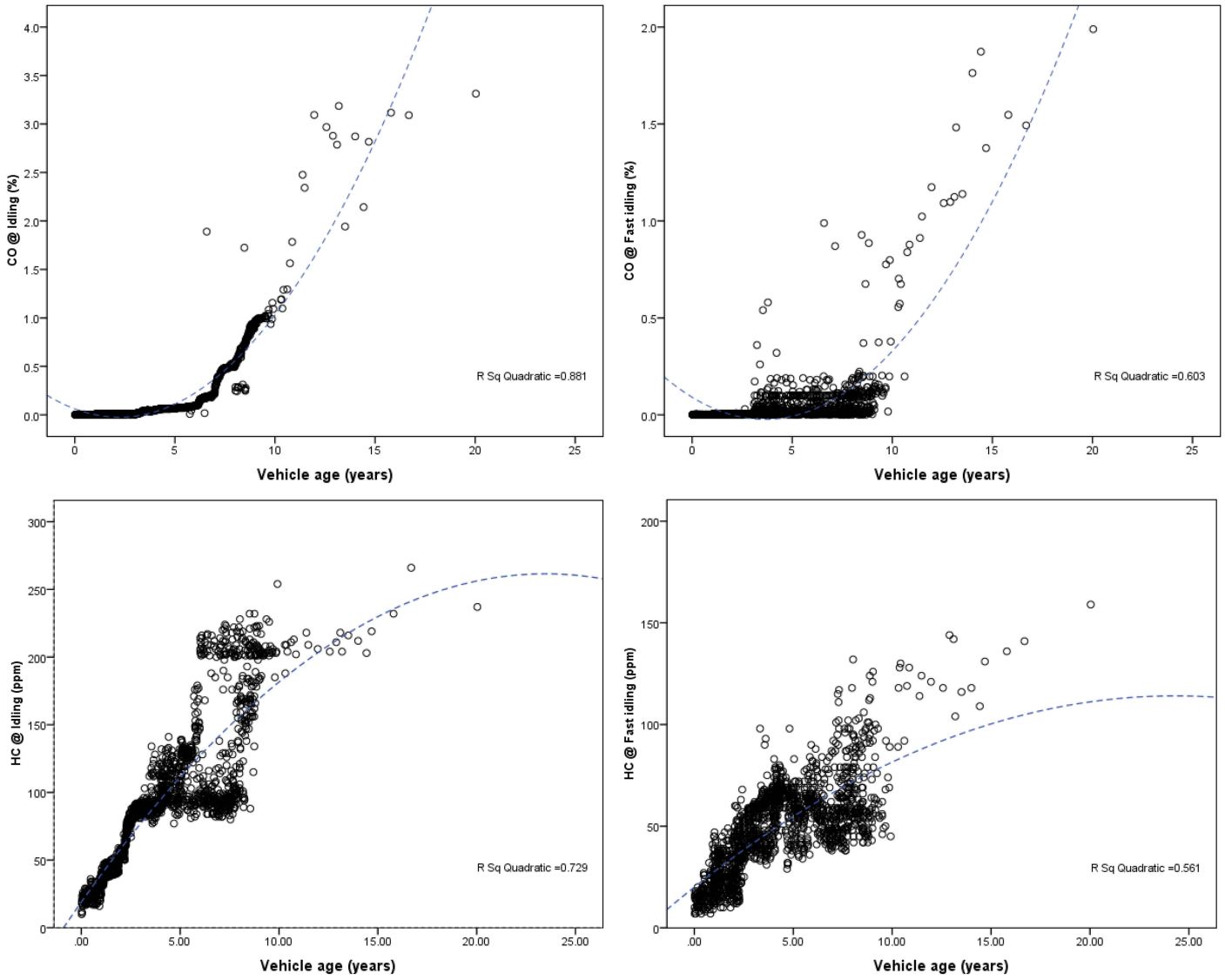


Fig. 2. Vehicle age versus CO and HC emissions.

J. Environ. Eng., 2022, 148(7): 04022035

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

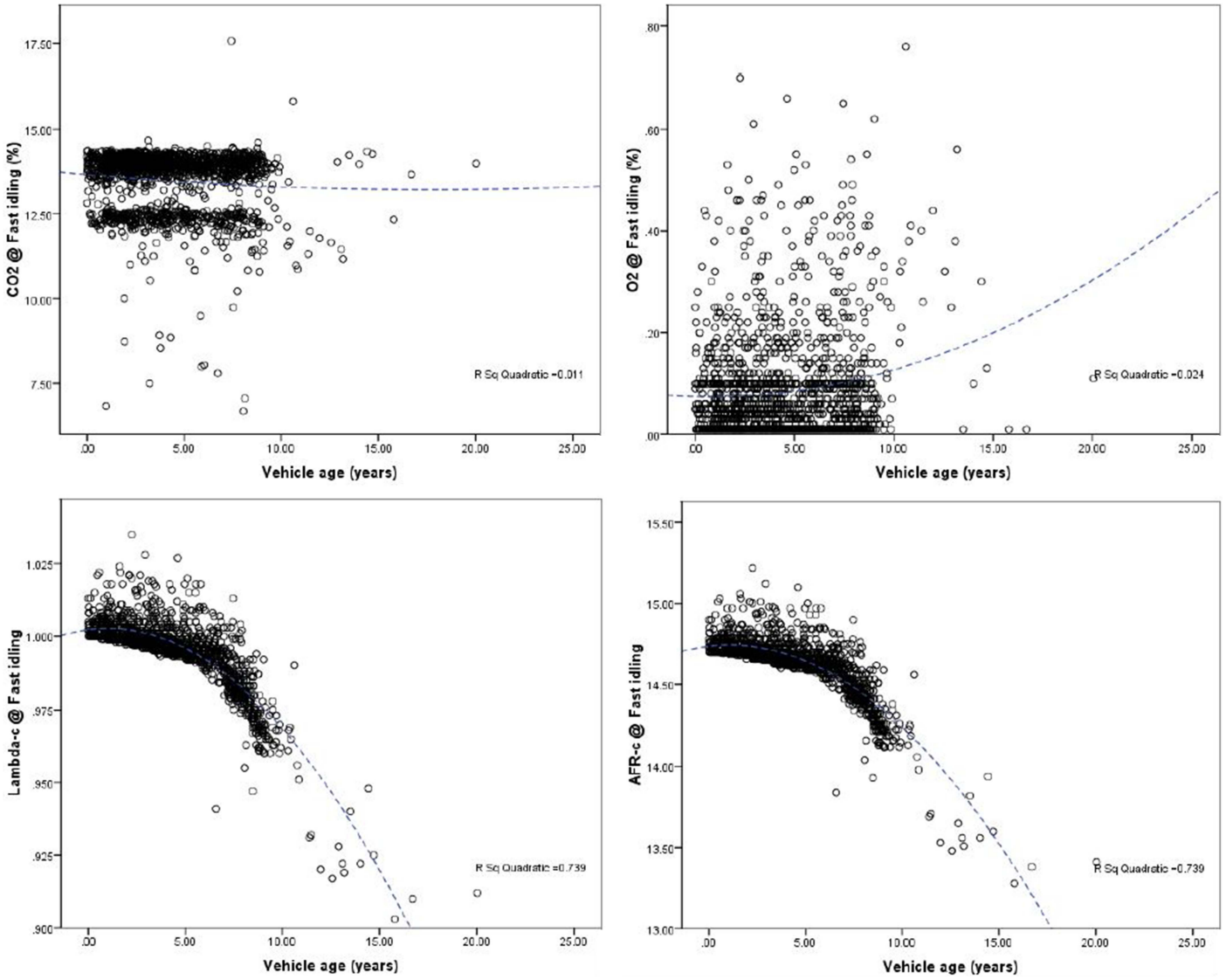


Fig. 3. Vehicle age versus CO_2, O_2 emissions, λ and AFR.

and maintenance category based on the number of fitness visits to the manufacturer's authorized service centers.

Further, the effects of engine-related aspects, i.e., cubic capacity, maximum power, maximum torque, compression ratio, cylinder bore, piston stroke, aspiration type, no. of cylinders, no. of valves per cylinder, valve configuration, and fuel mixing conditions are also plotted. For better resolution and understanding of the vehicle's performance towards tailpipe exhaust emission, the data analysis was carried out for the top 3 makes, representing 83.61% of the sample size (n = 1,580).

The average characteristics or range of the variables collected during the emissions testing program for petrol cars are presented in Table 3. All car makes studied had four-stroke engines with a spark ignition system. All cars were found to be fitted with 3-way catalytic converters. The tightening in registration norms (maximum 15 years for petrol cars from the data of registration) was reflected with only 3 cars violating this norm. The petrol cars tested were usually younger (average age <4.5 years). MSIL had the largest chunk of the market share with its models also leading the sales in the relevant vehicle class.

Vehicle Age

The effect of vehicle age on CO and HC emissions for the entire data range in both idle and fast idle testing modes is shown in Fig. 2. It is revealed that the emission characteristics for both parameters are best described by a 2nd order polynomial curve even though some scatter in the data is conspicuous. The effect of vehicle age on other tailpipe emission parameters, i.e., CO_2, O_2, λ, and AFR was also plotted. CO_2 and O_2 emissions along with λ and

Table 4. Emission equations with respect to vehicle age (whole dataset)

Parameter	Equation	R^2
CO at idling	$y = 0.0174x^2 - 0.0772x + 0.0615$	0.881
CO at fast idling	$y = 0.0087x^2 - 0.063x + 0.0908$	0.603
HC at idling	$y = -0.4401x^2 + 20.663x + 18.968$	0.729
HC at fast idling	$y = -0.1592x^2 + 7.739x + 20.058$	0.561
CO_2 at fast idling	$y = 0.0014x^2 - 0.0509x + 13.661$	0.011
O_2 at fast idling	$y = 0.0006x^2 - 0.0009x + 0.0743$	0.024
λ at fast idling	$y = -0.0004x^2 + 0.0009x + 1.0022$	0.739
AFR at fast idling	$y = -0.0063x^2 + 0.0134x + 14.736$	0.738

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

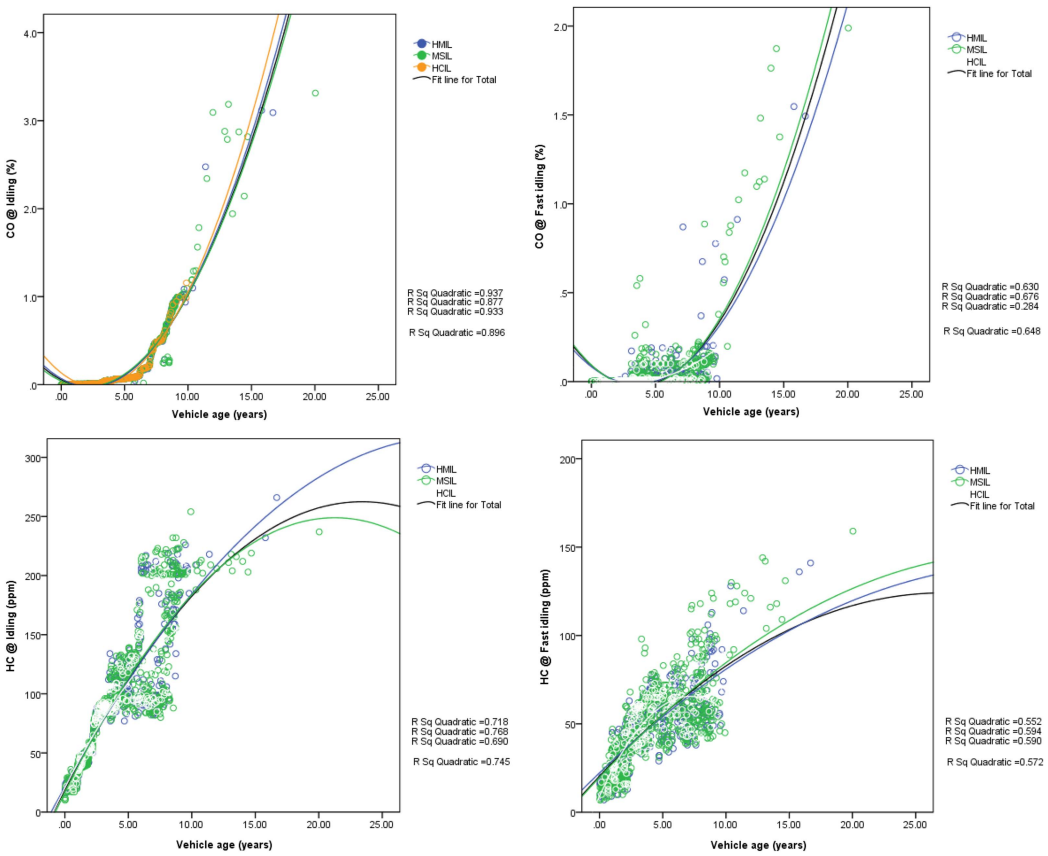


Fig. 4. Vehicle age versus CO and HC emissions for top 3 makes.

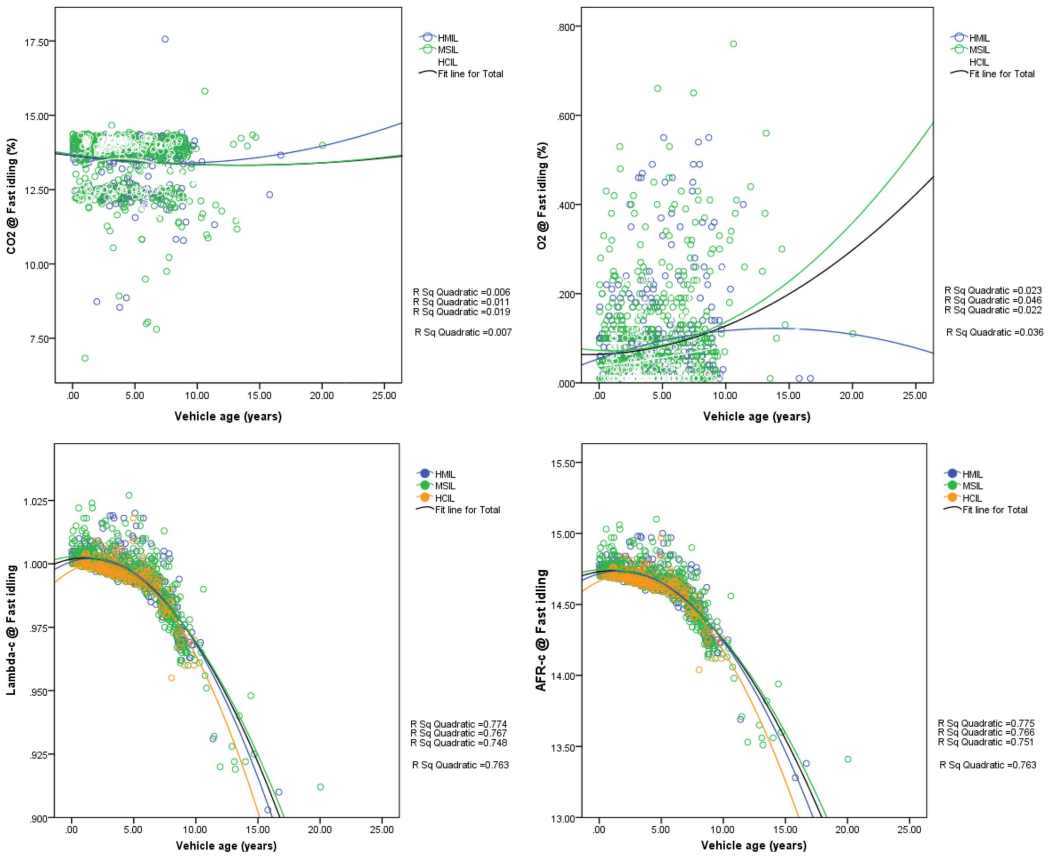


Fig. 5. Vehicle age versus CO_2, O_2 emissions, λ and AFR for top 3 makes.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Table 5. Emission equations with respect to vehicle age (top 3 makes)

Parameter	MSIL		HMIL		HCIL	
	Equation	R^2	Equation	R^2	Equation	R^2
CO at idling	$y = 0.0163x^2 - 0.0655x + 0.0404$	0.877	$y = 0.0178x^2 - 0.0821x + 0.0696$	0.937	$y = 0.0205x^2 - 0.1127x + 0.1339$	0.933
CO at fast idling	$y = 0.0094x^2 - 0.0678x + 0.0935$	0.676	$y = 0.0079x^2 - 0.0559x + 0.0838$	0.631	$y = 0.0041x^2 - 0.0254x + 0.0377$	0.284
HC at idling	$y = -0.5139x^2 + 21.854x + 16.574$	0.768	$y = -0.322x^2 + 19.525x + 21.428$	0.718	$y = -0.7055x^2 + 21.364x + 20.595$	0.692
HC at fast idling	$y = -0.1144x^2 + 7.6276x + 19.781$	0.594	$y = -0.0991x^2 + 6.8566x + 22.266$	0.552	$y = -1.0946x^2 + 15.711x + 9.1707$	0.592
CO_2 at fast idling	$y = 0.002x^2 - 0.0549x + 13.69$	0.011	$y = 0.0046x^2 - 0.066x + 13.637$	0.006	$y = -0.0234x^2 + 0.1849x + 13.179$	0.019
O_2 at fast idling	$y = 0.0008x^2 - 0.0016x + 0.0728$	0.046	$y = -0.0004x^2 + 0.0098x + 0.0545$	0.023	$y = 0.0001x^2 + 0.0032x + 0.0375$	0.022
λ at fast idling	$y = -0.0004x^2 + 0.0004x + 1.0029$	0.767	$y = -0.0005x^2 + 0.0016x + 1.0009$	0.774	$y = -0.0006x^2 + 0.0027x + 0.9974$	0.748
AFR at fast idling	$y = -0.0055x^2 + 0.0065x + 14.745$	0.766	$y = -0.0072x^2 + 0.0238x + 14.716$	0.775	$y = -0.009x^2 + 0.0405x + 14.664$	0.751

AFR for the entire data range are shown in Fig. 3. Further, λ and AFR show a strong correlation with vehicle age, implying that both reduce with aging vehicles.

On the other hand, the other two parameters, i.e., CO_2 and O_2 emissions, were found to be not related to vehicle age with R^2 values that were too low, pointing at no direct correlation. The emission equations along with corresponding R^2 values computed using the 2nd order polynomial curve presenting the likelihood of correlation amongst all eight parameters with respect to vehicle age for idle and fast idle test methods are given in Table 4.

The field-measured tailpipe emission parameters with respect to vehicle age were also plotted for the top 3 makes of petrol-driven passenger cars, based on their representativeness in the sample size. This additional analysis aimed at better data resolution and variability amongst makes apart from the whole data. The data was analyzed for idle and fast idle test modes, and the results are shown in Figs. 4 and 5 for the top 3 makes. Again, it was observed that the dependence of emission levels on vehicle age is polynomial or quadratic, in nature. The resultant emission equations and R^2 values obtained for the best fit quadratic trendlines in the case of the top 3 makes are presented in Table 5.

A glance at emission equations having vehicle age as an independent variable reveals that the aging of cars directly influences the emission characteristics. This may lead to non-compliance of vehicles towards concurrent emission norms thereby necessitating a quick (in the form of engine tuning or replacement of the catalytic converter, etc.) or long-haul inspection (in the form of maintenance of the whole vehicle). However, significant data scattering may be observed in fast idling conditions for both CO and HC (Fig. 2), resulting in relatively poorer correlation values ($R^2 = 0.603$ and 0.561, respectively). This might be because high idle conditions (high engine RPM) are difficult to be idealized and, in many cases, are accompanied by abrupt variations in engine speed and, therefore, the unsteady concentrations encountered in the corresponding readings.

The emissions equations further computed for the top 3 makes are generally in line with those for the whole data set except for HCIL make, which shows a poor or no correlation when comparing vehicle age versus fast idling CO ($R^2 = 0.284$). The study further indicated that CO_2 and O_2 exhaust concentrations do not seem to be affected by the vehicle age for fast idle conditions for both cases, i.e., entire dataset and top 3 makes, whereas λ and AFR at fast idling were found to be fairly correlated with vehicle age (Fig. 3).

Vehicle Mileage

Figs. 6 and 7 present the plots drawn to understand how vehicle mileage is related to tailpipe parameters in both idle and fast idle testing modes for the entire dataset. As can be seen, a 2nd order polynomial curve best describes the emission characteristics for both pollutants, even if the data scatter is obvious. The resultant R^2 values indicate how the entire dataset behaved in the scatterplots giving a conclusive picture. It was found that vehicle mileage is significantly and positively related to CO and HC and negatively correlated to λ and AFR, implying that vehicles with more accumulated mileage show increased values of CO and HC while reduced values for both λ and AFR, respectively. On the other hand, no correlation could be established between the other two parameters, i.e., CO_2 and O_2 emissions, with respect to vehicle mileage with R^2 values that were too low.

The emission equations and corresponding R^2 values computed using the 2nd order polynomial curve presenting the likelihood of correlation amongst all eight parameters with respect to vehicle

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

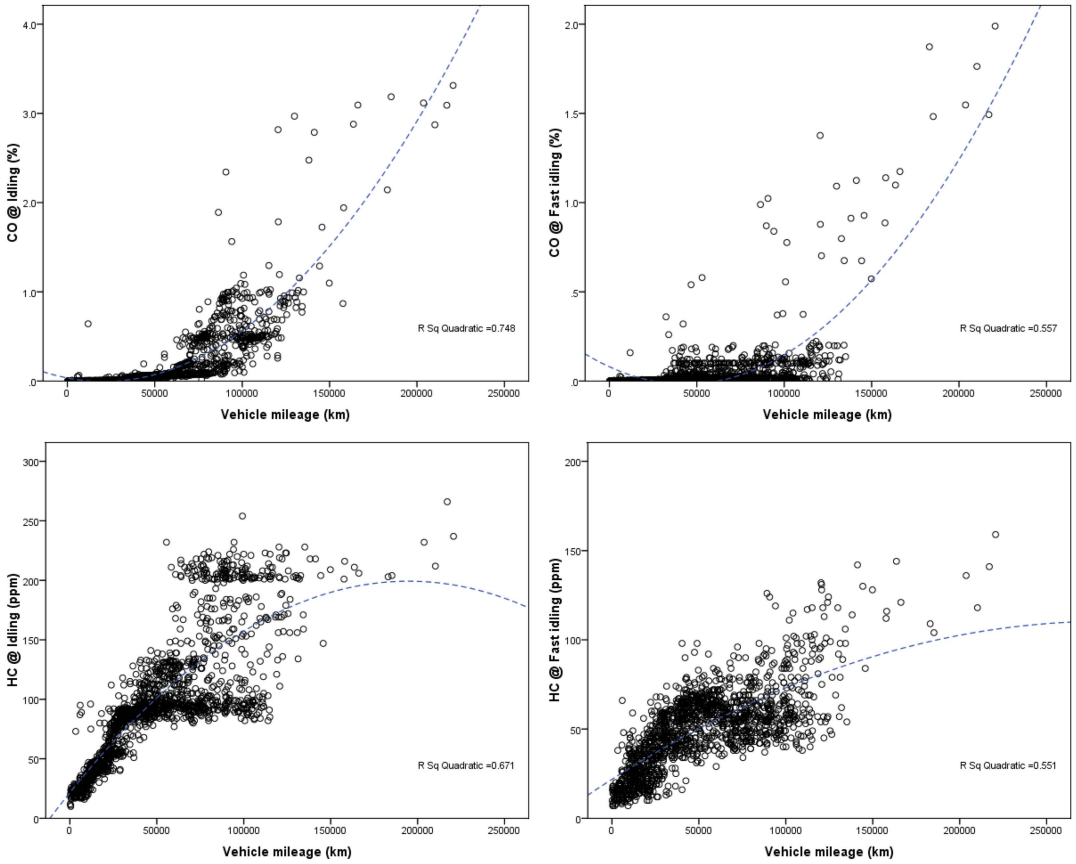


Fig. 6. Vehicle mileage versus CO and HC emissions.

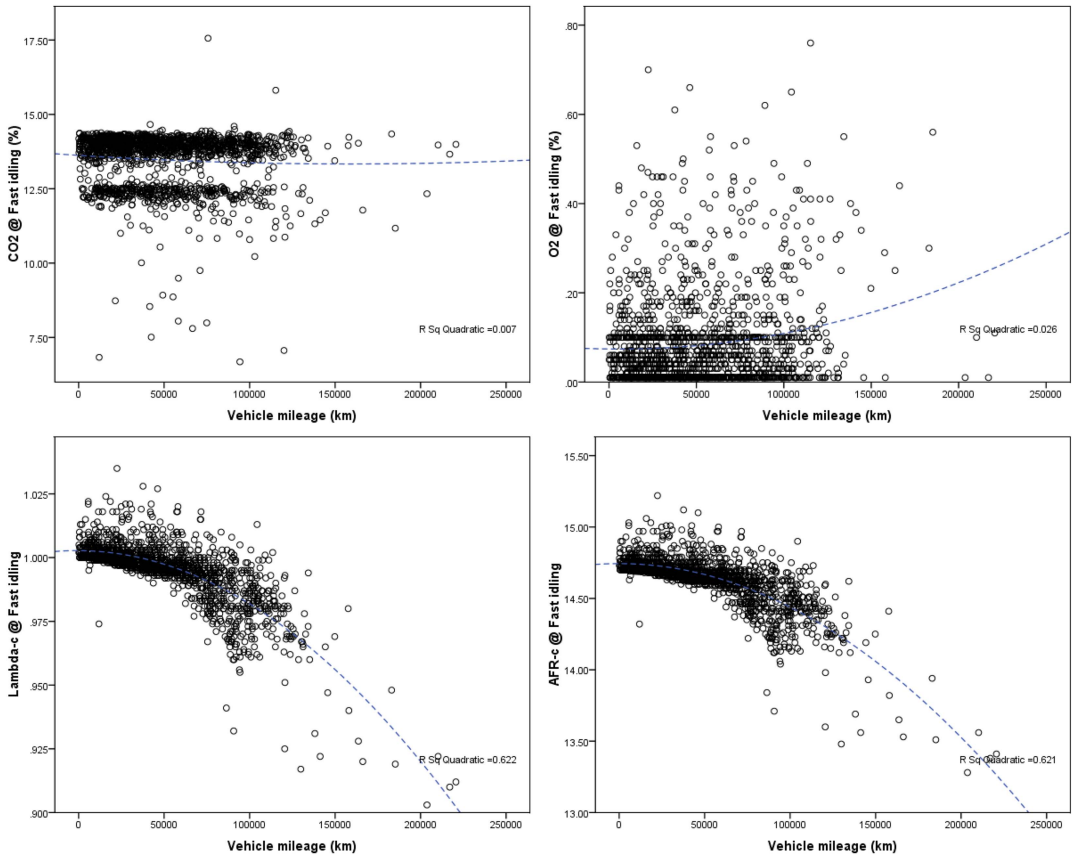


Fig. 7. Vehicle mileage versus CO_2, O_2 emissions, λ and AFR.

Table 6. Emission equations with respect to vehicle mileage (whole dataset)

Parameter	Equation	R^2
CO at idling	$y = 0.0000000001x^2 - 0.0000037836x + 0.0346$	0.748
CO at fast idling	$y = 0.0000000001x^2 - 0.0000044730x + 0.0801$	0.557
HC at idling	$y = -0.0000000047x^2 + 0.0018247565x + 21.6145$	0.671
HC at fast idling	$y = -0.0000000011x^2 + 0.0006223787x + 21.8187$	0.551
CO_2 at fast idling	$y = 0.00000000001x^2 - 0.00000366929x + 13.6214$	0.007
O_2 at fast idling	$y = 0.000000000004x^2 - 0.000000045484x + 0.0741$	0.026
λ at fast idling	$y = -0.000000000002x^2 - 0.000000004131x + 1.0027$	0.622
AFR at fast idling	$y = -0.00000000003x^2 - 0.00000000697x + 14.7436$	0.621

mileage for idle and fast idle test methods are given in Table 6 (whole dataset) and Table 7 (Top 3 make-wise).

The measured concentration data points for tailpipe emission parameters, vis-à-vis vehicle mileage, were also plotted for the top 3 makes of passenger cars under investigation (Figs. 8 and 9). Again, the dependence of emission levels or parameters on vehicle mileage is quadratic in nature, giving the best fit for trendlines. The resultant emission equations and R^2 values obtained from such quadratic trendlines in the case of top 3 makes are shown in Table 8.

A glance at emission equations having vehicle mileage as an independent variable finds that the accumulated kilometers of cars directly impact the emission characteristics or parameters. Higher vehicle mileage over a period of time may also make a vehicle non-compliant towards extant emission norms. This non-compliance may need to be addressed either by way of a quick-fix or thorough maintenance in a bid to bring the vehicle back to compliance. Although data scattering may be observed in fast idle conditions for both CO and HC (Figs. 6 and 8), resulting in poor correlation values (R^2) ranging between 0.5 and 0.6, it is slightly better than those reported in the case of vehicle age.

The emission equations calculated for the top 3 makes, as given in Table 8, are generally in line with those for the whole data set except for HCIL, which showed a poor or no correlation when comparing vehicle mileage versus fast idle CO ($R^2 = 0.295$). This scenario is the same as found in the case of vehicle age. The study further indicated that CO_2 and O_2 exhaust concentrations do not seem to be affected by the vehicle mileage for fast idling conditions for both analysis cases, i.e., entire data set and top 3 makes, whereas λ and AFR at fast idling were found to be fairly correlated with vehicle mileage (Figs. 7 and 9).

An overview of the plots relating vehicle age and mileage also indicates a comparison of the trendlines for the top 3 makes based on their representation in the entire data set. Assuming that trendlines are the guiding factor, the relative standing of different makes in terms of their CO and HC emissions (in both testing modes) due to aging and (accumulated) mileage effects in descending order is presented in Table 8.

Vehicle Emission Norm (at Manufacturing Time)

Exhaust emission norms are progressively becoming more stringent worldwide with a primary aim of inducing the production of vehicles having better performance towards compliance to emission

Table 7. Emission equations with respect to vehicle mileage (top 3 makes)

	MSIL		HMIL		HCIL	
Parameter	Equation	R^2	Equation	R^2	Equation	R^2
CO at idling	$y = 0.0000000001x^2 - 0.0000041163x + 0.0393$	0.755	$y = 0.0000000001x^2 - 0.0000032311x + 0.0233$	0.821	$y = 0.0000000001x^2 - 0.0000033536x + 0.0338$	0.749
CO at fast idling	$y = 0.0000000001x^2 - 0.0000052871x + 0.0923$	0.616	$y = 0.0000000004x^2 - 0.0000361277x + 0.0668$	0.622	$y = 0.0000000002x^2 - 0.0000154364x + 0.0285$	0.295
HC at idling	$y = -0.0000000005x^2 + 0.0019189444x + 18.7679$	0.716	$y = -0.0000000004x^2 + 0.001665741x + 26.9806$	0.648	$y = -0.0000000008x^2 + 0.002013920x + 21.1361$	0.605
HC at fast idling	$y = -0.0000000006x^2 + 0.0011418895x + 14.0273$	0.592	$y = -0.0000000002x^2 + 0.0004876134x + 25.7874$	0.551	$y = -0.0000000009x^2 + 0.0006295487x + 20.9631$	0.551
CO_2 at fast idling	$y = 0.0000000002x^2 - 0.00000385452x + 13.6335$	0.004	$y = 0.0000000001x^2 - 0.0000345688x + 13.5927$	0.004	$y = -0.0000000001x^2 + 0.0000111342x + 13.2564$	0.017
O_2 at fast idling	$y = 0.0000000000006x^2 - 0.000000212153x + 0.0737$	0.058	$y = -0.0000000000002x^2 + 0.000000725923x + 0.0587$	0.021	$y = 0.0000000000003x^2 + 0.0000002462571x + 0.0405$	0.014
λ at fast idling	$y = -0.0000000000002x^2 - 0.000000009572x + 1.0027$	0.639	$y = -0.0000000000002x^2 + 0.000000025717x + 1.0021$	0.676	$y = -0.0000000000002x^2 + 0.000000026046x + 1.0005$	0.617
AFR at fast idling	$y = -0.0000000000003x^2 - 0.000000008273x + 14.7430$	0.638	$y = -0.0000000000003x^2 + 0.000000040785x + 14.7341$	0.675	$y = -0.0000000000003x^2 + 0.000000047583x + 14.7105$	0.618

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

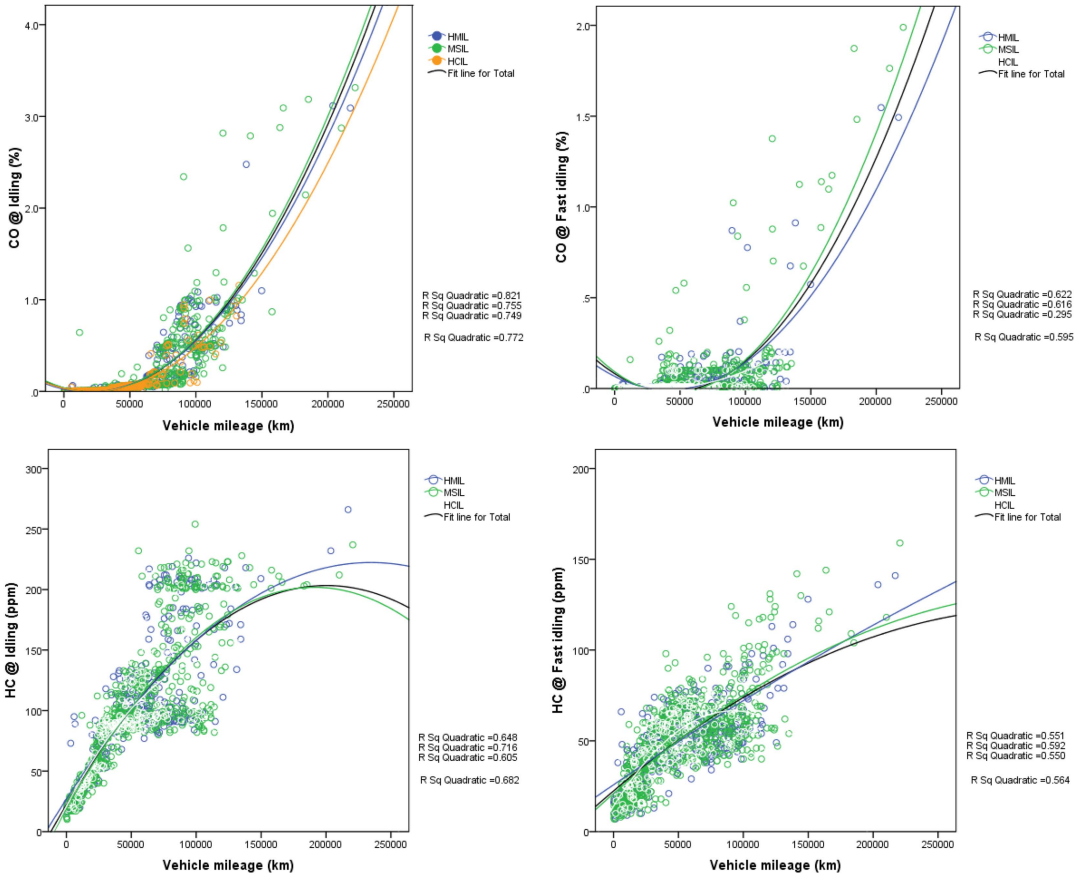


Fig. 8. Vehicle mileage versus CO and HC emissions for top 3 makes.

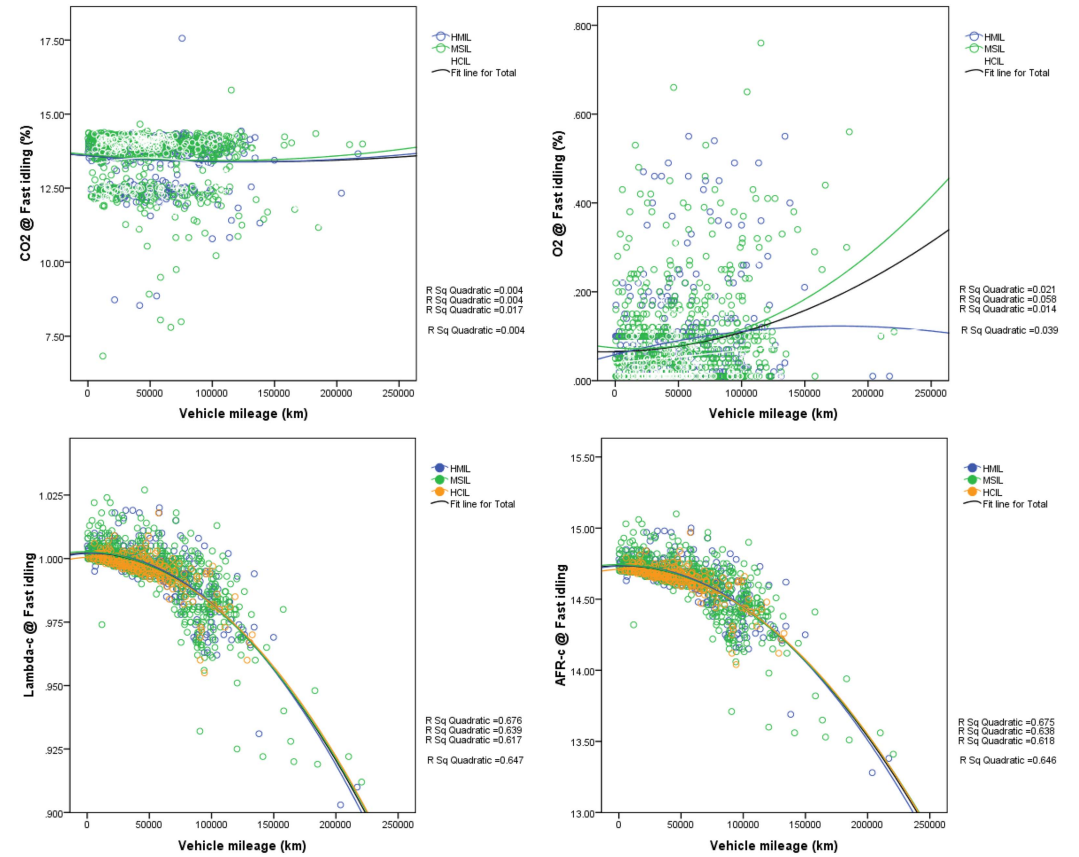


Fig. 9. Vehicle mileage versus CO_2, O_2 emissions, λ and AFR for top 3 makes.

Table 8. Relative standing of different makes in descending order in terms of their CO and HC emissions due to age and mileage factor

S. no.	Aging effect on CO emission		Aging effect on HC emission	Mileage effect on CO emission		Mileage effect on HC emission
	Idle	Fast idle	Idle	Idle	Fast idle	Idle
1.	HCIL	MSIL	HMIL	MSIL	MSIL	HMIL
2.	HMIL	HMIL	MSIL	HMIL	HMIL	MSIL
3.	MSIL	HCIL	HCIL	HCIL	HCIL	HCIL

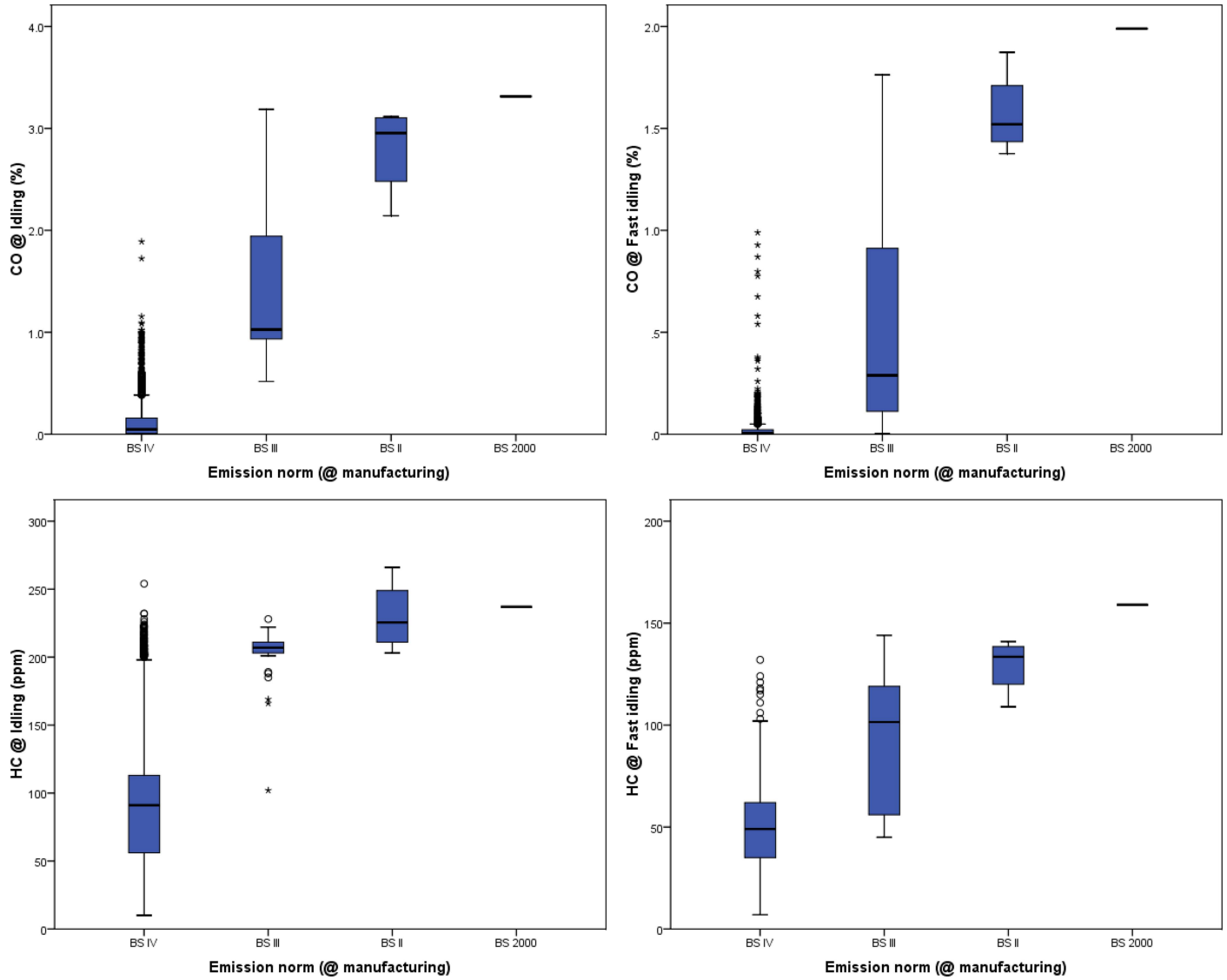


Fig. 10. Emission norm versus CO and HC emissions.

norms thereby achieving a reduction in automobile-produced air pollution. In this connection, the current study investigated the effect of concurrent emission norms on tailpipe emission and other related parameters. The analyzed emission norms were BS (Bharat Stage) norms (such as BS 2000, BS II, BS III, and BS IV) in India which were adopted in line with Euro norms. For the whole dataset, the CO, HC (under both idle and fast idle testing conditions) and O_2 emissions (under fast idling conditions) were found to improve for a relatively newer norm. The range of emission of these parameters was found to decrease with the introduction of stringent norms following the order BS IV > BS III > BS II > BS 2000 (introduced in India in the years 2010, 2005, 2001, and 2000, respectively).

Only CO_2 emissions were found to be independent of any emissions norm and did not fluctuate significantly with the onset of any BS norm. Two other parameters, λ and AFR, were also reported to be strongly influenced by the emission norms and depicted reduction in ranges with the application of newer norms in the same manner as in CO, HC, and O_2 emissions (Fig. 10 and Fig. S3). All vehicle makes (HMIL, MSIL, and HCIL) also report a considerable degree of reduction in CO, HC, and O_2 emission levels between vehicles having BS III as the complying emission norm compared to those with BS IV. BS IV has a remarkably lower range of CO, HC, and O_2 emissions and has been able to bring about a significant reduction in CO and HC levels in both test conditions.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

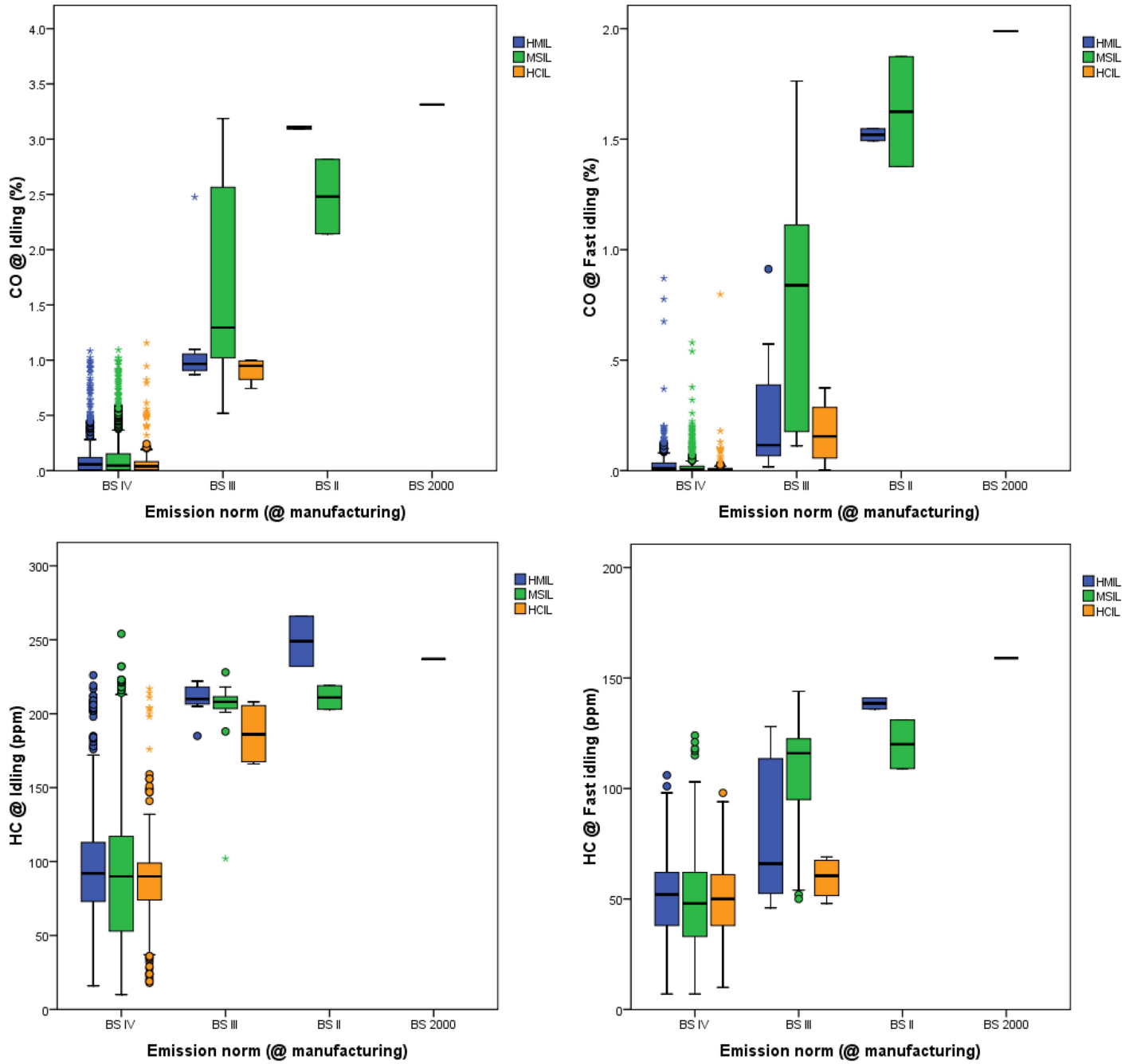


Fig. 11. Emission norm versus CO and HC emissions for top 3 makes.

BS III also presents a reasonable degree of CO and HC emissions reduction compared to BS II. It is seen that the MSIL and HMIL make have the sharpest reduction in emissions levels, while HCIL seemed to follow a somehow lesser variation (Fig. 11 and Fig. S4). λ, AFR, CO_2, and O_2 followed the trend as observed in the case of the whole dataset.

Vehicle Maintenance Category

The effect of regular I/M has been reported to be directly related to a vehicle's emission levels. Vehicles with a poor I/M history tend to emit more CO and HC under both idle and fast idle testing conditions and vice versa. The present study aims to reinvestigate the validity of this finding with a larger dataset, a few more tailpipe parameters, and different makes. The criteria to categorize vehicles based on their number of I/M visits paid to the respective authorized service centers for periodic/casual fitness checks over the last two years was used and is presented in Table 9. This criterion considers a frequency of every two years for vehicles approaching their respective manufacturer's authorized service centers for fitness tests.

Boxplots were drawn showing the effect of such maintenance categories on tailpipe parameters. It was revealed that the maintenance level has a very apparent and negative correlation with CO and HC emissions under idle and fast idle conditions. The better the maintenance level or higher the number of fitness visits to the authorized service centers, the lower the CO and HC levels in the tailpipe exhaust. Although few outliers are visible, especially in

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Table 9. Criteria for categorization of vehicle's maintenance level

S. no.	No. of fitness visits to manufacture's authorized service centers	Vehicle's maintenance category
1	5–7	Very good
2	3–4	Good
3	1–2	Poor
4	0	Unsatisfactory

the case of the 'unsatisfactory' and 'good' maintenance category or class of vehicles, and more so during fast idling, they do not seem to adversely affect the strong degree of overall lower ranges of emission, reducing with betterment in vehicle maintenance category (Fig. 12). While CO_2 and O_2 emissions are found to be unaffected by maintenance category depicting unchanging ranges, λ and AFR were seen to decrease with a higher number of fitness visits (Fig. S5).

In terms of make, HCIL depicted the lowest range of CO and HC emissions and the quickest reduction in concentration ranges in both testing conditions with respect to better vehicle maintenance records, followed by MSIL; however, the maximum inconsistency was observed in the case of HMIL, especially in HC in the fast idling scenario (Fig. 13). CO_2 and O_2 emissions and λ and AFR values were found to be in line with that depicted in the case of the whole dataset (Fig. S6).

Effect of Engine-Related Variables on Tailpipe Emission/Parameters

The study also attempted to investigate the effect(s) of various engine-specific variables (scale and/or strings type datapoints) on tailpipe emission parameters, such as engine capacity (cc), maximum power (bhp), maximum torque (Nm), compression ratio (:1), cylinder bore (mm) and piston stroke (mm), engine aspiration type (natural or turbo-charged), number of cylinders, number of valves per cylinder, valve configuration (SOHC–single overhead cam or DOHC–double overhead cam) and fuel mixing conditions (rich, lean, or stoichiometric). It is revealed that engine variables, namely, engine capacity, maximum power, maximum torque, cylinder bore and piston stroke, number of cylinders, number of valves

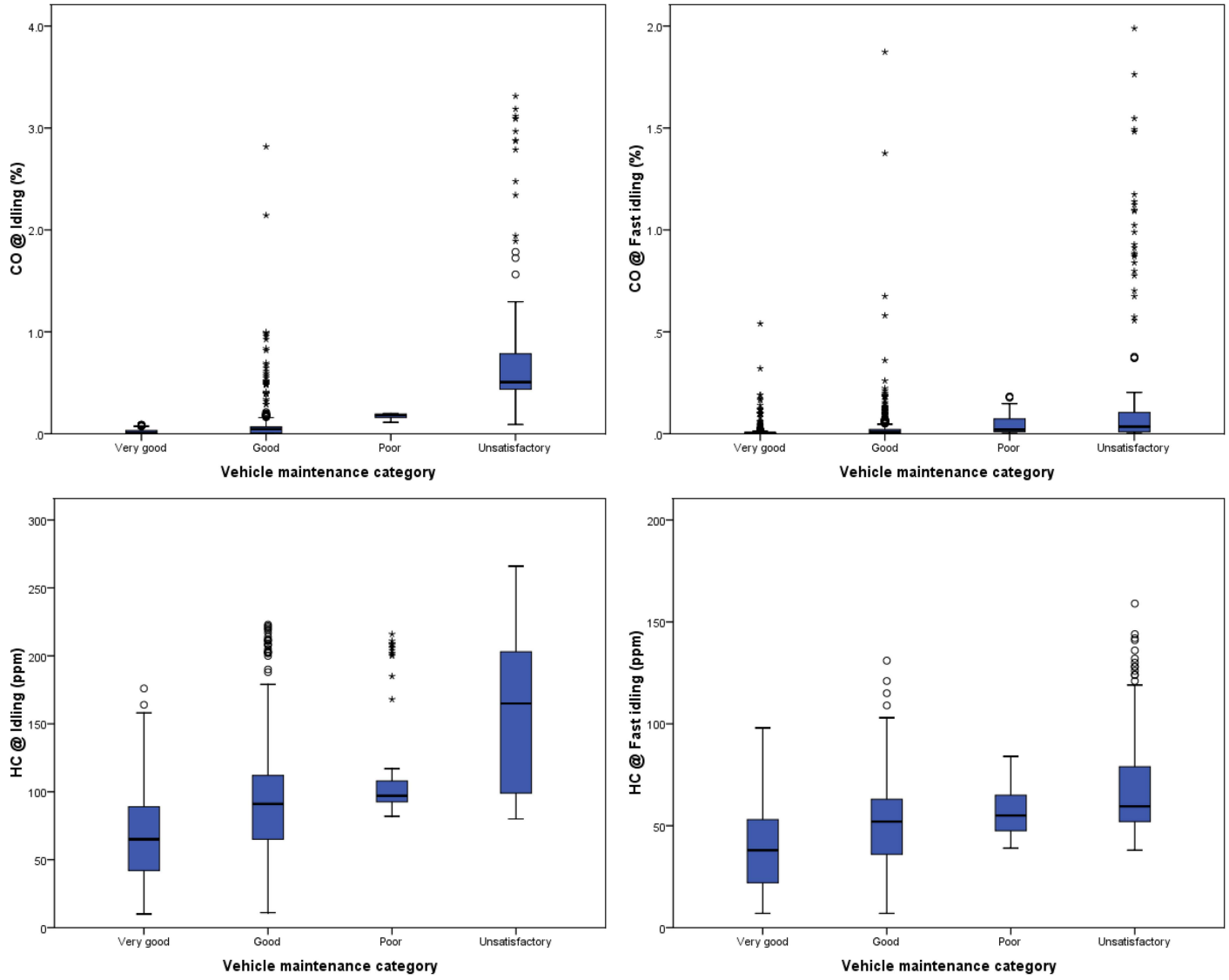


Fig. 12. Vehicle maintenance versus CO and HC emissions.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

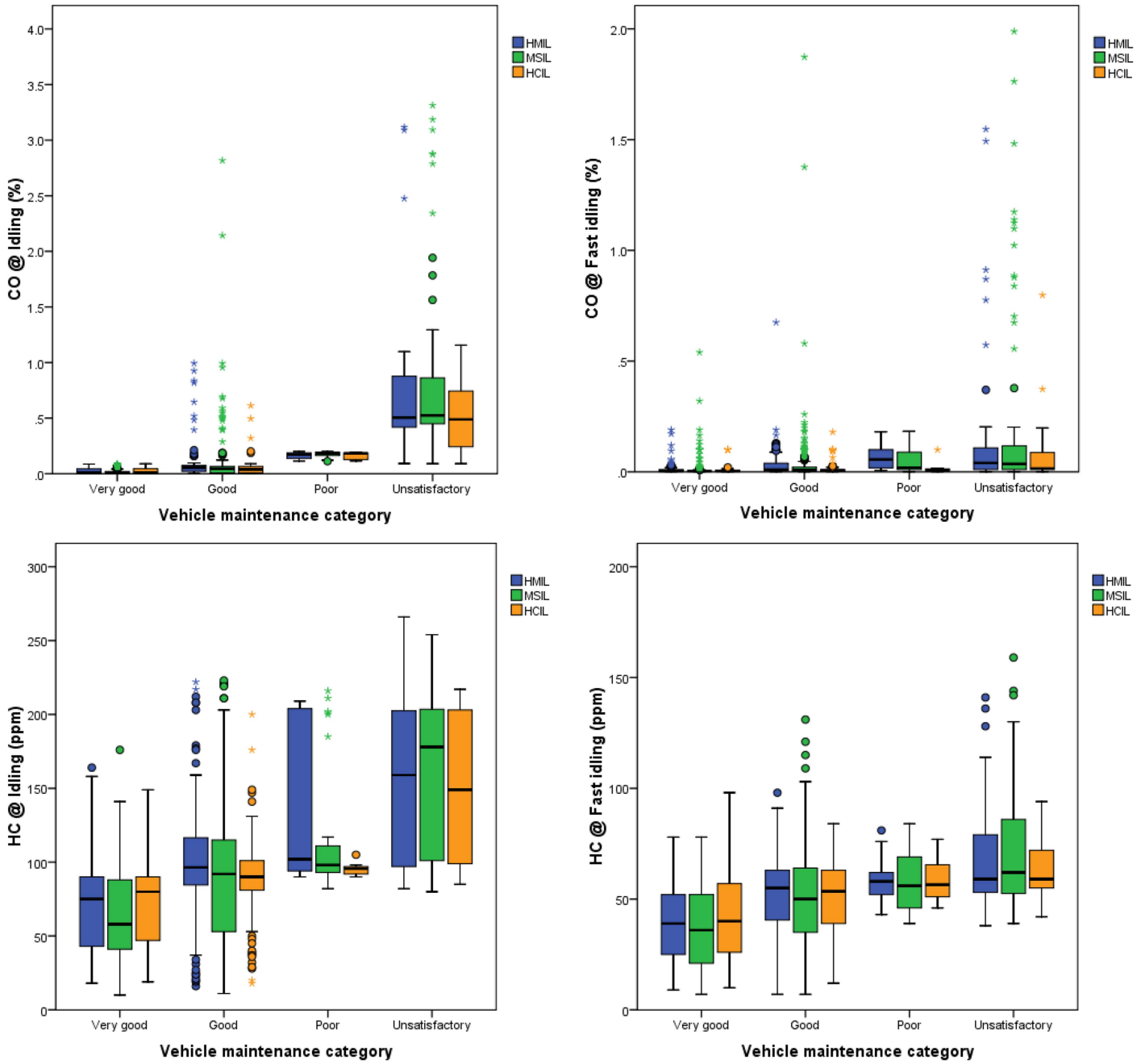


Fig. 13. Vehicle maintenance versus CO and HC emissions for top 3 makes.

per cylinder, and valve configuration do not seem to be related to tailpipe exhaust (CO, HC, CO_2, and O_2) or other parameters, such as λ and AFR in both dataset scenarios.

On the other hand, two (2) engine-related independent variables, i.e., engine aspiration type and fuel mixing conditions, were found to be related with the tailpipe parameters for both dataset conditions. Turbo-charged engines were found to emit lower ranges of both CO and HC under both idle and fast idle testing modes compared to naturally aspirated ones and both datasets had similar findings (Fig. 14, Fig. S8).

The fuel mixing conditions also reported a good relationship with tailpipe emission. CO and HC had the lowest ranges of emission of CO and HC in idling and fast idling test modes during stoichiometric fuel mixing or burning condition, relatively higher during lean, and the highest during rich fuel burning scenarios prevalent in the engine's combustion chamber. This finding was found to be consistent with the concentration ranges or values of CO_2, O_2, λ, and AFR, as revealed by the data analysis (Fig. S10). Furthermore, the make-wise scenario was also found to be similar as in the case of the whole dataset for CO and HC (Fig. 15) and for other tailpipe parameters (Fig. S11).

Conclusions

The present study catered to tailpipe emission monitoring of 1,580 petrol-driven passenger cars in the NCT of Delhi.

The study reported better correlation values relating age and mileage as prominent vehicular aspects affecting tailpipe parameters

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

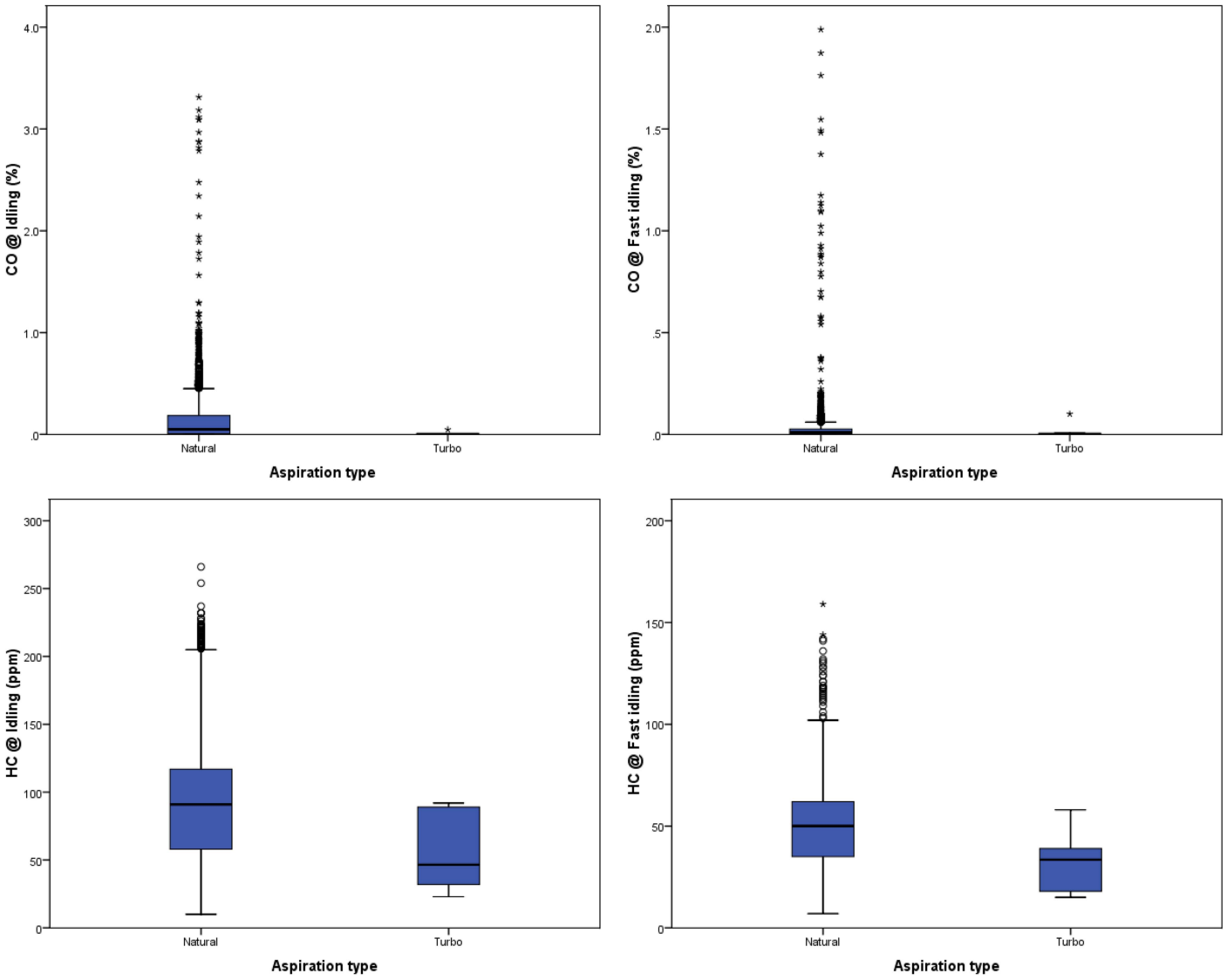


Fig. 14. Engine aspiration versus CO and HC emissions.

(CO, HC, λ, and AFR versus age, $R^2 = 0.881$, 0.729, 0.739, and 0.738, respectively, and versus mileage, $R^2 = 0.748$, 0.671, 0.622, and 0.621). Make-wise scenarios even bettered such findings for top 3 petrol-driven car makes (R^2 values for MSIL considering age and CO, HC, λ, and AFR stood as 0.877, 0.768, 0.767, and 0.766; for HMIL as 0.937, 0.718, 0.774, and 0.775; and for HCIL as 0.933, 0.692, 0.748, and 0.751, respectively, while considering mileage, R^2 values for MSIL fared as 0.755, 0.716, 0.639, and 0.638; for HMIL as 0.821, 0.648, 0.676, and 0.675; and for HCIL as 0.749, 0.605, 0.617, and 0.618, respectively). Vehicle emission norms and maintenance categories lined-up as another significant aspect(s) affecting tailpipe emission parameters.

We report that due to the country's lenient idle testing certification norms, low failure rates were observed, which could be a deviation from the actual emissions status of passenger cars. This scenario is evident in most developing countries where I/M programs are fundamental and need to be suitably revised to link high idle testing for PUC certification while also looking into lowering emission norms (Fig. 16). It is also stated that the CO should be given higher weighting in emissions recertification than HC, as the latter has a poorer correlation with aging cars (specifically in idle testing mode, Table 10).

We have computed the emission equations that can be used to forecast the emission values of CO, HC, and other parameters for the whole dataset and also for top 3 makes using the vehicle age and mileage without a need to actually test them at the PUC or any other test centers. As part of checking the emissions estimation accuracy using emissions equations, we find that CO and HC emission values as predicted by the corresponding equations are strongly correlated to the actual measured values (R^2 for CO = 0.94, and for HC = 0.93) amid CO faring slightly better than HC in terms of correlation values alone (Figs. S13 and S15). Moreover, there exists a good degree of agreement between the CO and HC emission values as predicted by the emission equations versus values measured in-situ.

Outcomes relative to top 3 makes may help the manufacturers adopt a suitable strategy for tuning the engine or vehicle before they become non-compliant to emission norms or bring them back to compliance levels. As an essential policy measure, the study recommends that the governments and policymakers in developing countries should investigate phasing out the old and more highly polluting cars from the roads than tightening emission norms alone. As more stringent emission norms are being adopted worldwide, the scope of analyzing a model or fuel-specific emission dataset

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

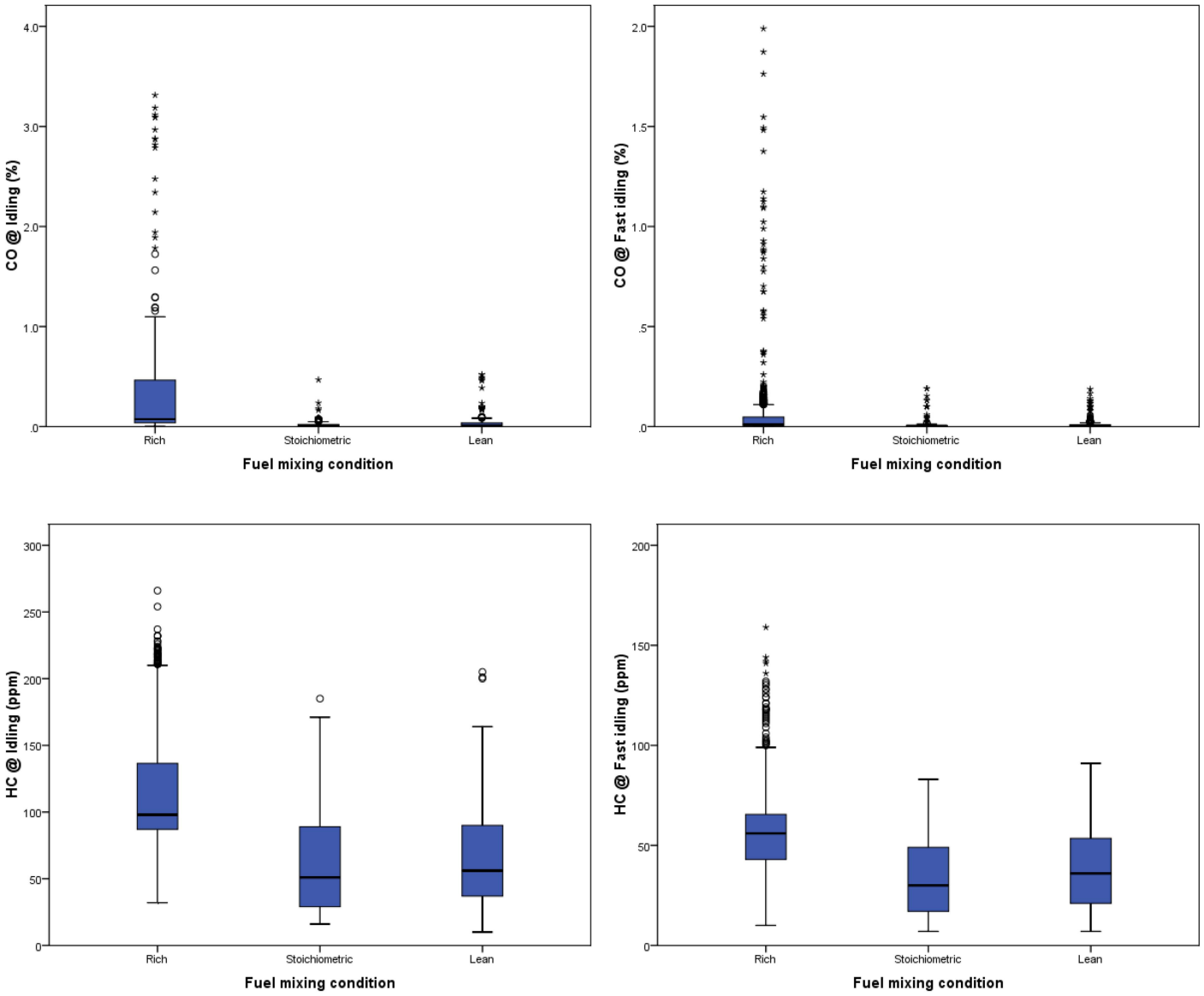


Fig. 15. Fuel mixing condition versus CO and HC emissions.

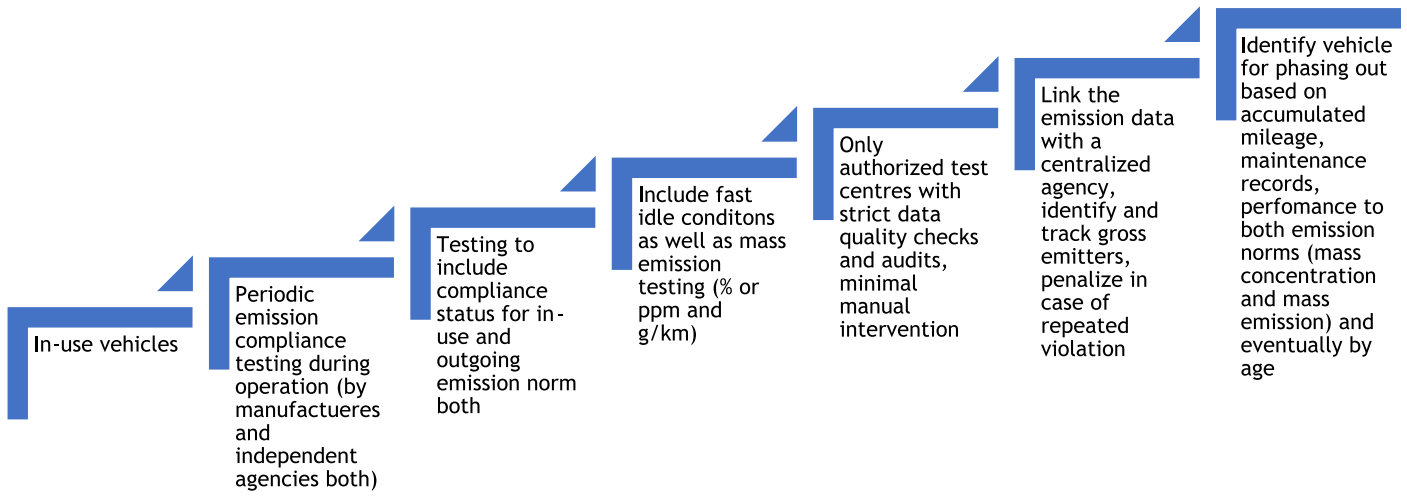


Fig. 16. Proposed emission compliance testing framework for in-use cars.

Table 10. Summary of correlation values for vehicle/engine-related variables and tailpipe parameters

Vehicle/engine-related variables	CO at idling	CO at fast idling	HC at idling	HC at fast idling	CO_2 at fast idling	O_2 at fast idling	λ at fast idling	AFR at fast idling
R^2 values for age (whole dataset; n = 1,580)	0.881	0.603	0.729	0.561	0.011	0.024	0.739	0.738
R^2 values for mileage (whole dataset; n = 1,580)	0.748	0.557	0.671	0.551	0.007	0.026	0.622	0.621
R^2 values for age (MSIL; n = 735)	0.877	0.676	0.768	0.594	0.011	0.046	0.767	0.766
R^2 values for mileage (MSIL; n = 735)	0.755	0.616	0.716	0.592	0.004	0.058	0.639	0.638
R^2 values for age (HMIL; n = 381)	0.937	0.631	0.718	0.552	0.006	0.023	0.774	0.775
R^2 values for mileage (HMIL; n = 381)	0.821	0.622	0.648	0.551	0.004	0.021	0.676	0.675
R^2 values for age (HCIL; n = 205)	0.933	0.284	0.692	0.592	0.019	0.022	0.748	0.751
R^2 values for mileage (HCIL; n = 205)	0.749	0.295	0.605	0.551	0.017	0.014	0.617	0.618

or the one generated using real-world driving conditions may need attention to reduce vehicular emissions in developing countries further.

Data Availability Statement

All data, models, and code generated or used during the study appear in the published article.

Acknowledgments

The authors would like to thank the PUC centers located across NCT of Delhi and the Advanced Air and Acoustics Research Lab, Delhi Technological University (DTU) for the support extended during the setup of the tailpipe exhaust monitoring station, data collection, and analysis.

Supplemental Materials

Figs. S1–S16 and Table S1 are available online in the ASCE Library (www.ascelibrary.org).

References

Al-Arkawazi, S. A. F. 2019. "Analyzing and predicting the relation between air–fuel ratio (AFR), lambda (λ) and the exhaust emissions percentages and values of gasoline-fueled vehicles using versatile and portable emissions measurement system tool." *SN Appl. Sci.* 1 (11): 1370. https://doi.org/10.1007/s42452-019-1392-5.

Badami, M. G. 2005. "Transport and urban air pollution in India." *Environ. Manage.* 36 (2): 195–204. https://doi.org/10.1007/s00267-004-0106-x.

Beydoun, M., and J. Guldmann. 2006. "Vehicle characteristics and emissions: Logit and regression analyses of I/M data from Massachusetts, Maryland and Illinois." *Transp. Res. D Transp. Environ.* 11 (1): 59–76. https://doi.org/10.1016/j.trd.2005.09.003.

Butler, T. 2013. "Urban and global effects of megacity pollution." In *Disposal of dangerous chemicals in urban areas and mega cities. NATO science for peace and security series C: Environmental security*, edited by I. Barnes and K. Rudziński. Dordrecht: Springer.

Caserini, S., C. Pastorello, P. Gaifami, and L. Ntziachristos. 2013. "Impact of the dropping activity with vehicle age on air pollutant emissions." *Atmos. Pollut. Res.* 4 (3): 282–289. https://doi.org/10.5094/APR.2013.031.

CPCB (Central Pollution Control Board). 2010. "Status of the vehicular pollution control programme in India." Accessed October 10, 2019. https://cpcb.nic.in/archivereport.php.

Dandapat, S., T. Ghosh, U. Shankar, S. Maitra, and B. Maitra. 2020. "A relook at the pollution certification of in-use vehicles in India and a way forward." *Asian Transp. Stud.* 6 (Jan): 100020. https://doi.org/10.1016/j.eastsj.2020.100020.

Frey, H. C., A. Unal, N. M. Rouphail, and J. D. Colyar. 2003. "On-road measurement of vehicle tailpipe emissions using a portable instrument." *J. Air Waste Manage. Assoc.* 53 (8): 992–1002. https://doi.org/10.1080/10473289.2003.10466245.

Government of Delhi. 2020. "Population projections India." Accessed April 17, 2020. https://nhm.gov.in/New_Updates_2018/Report_Population_Projection_2019.pdf.

Kagawa, S., K. Hubacek, K. Nansai, M. Kataoka, S. Managi, S. Suh, and Y. Kudoh. 2013. "Better cars or older cars?: Assessing CO_2 emission reduction potential of passenger vehicle replacement programs." *Glob. Environ. Change.* 23 (6): 1807–1818. https://doi.org/10.1016/j.gloenvcha.2013.07.023.

Kazopoulo, M., M. E. Fadel, and I. Kaysi. 2005. "Emission standards development for an inspection/maintenance program." *J. Environ. Eng.* 131 (9): 1330–1339. https://doi.org/10.1061/(ASCE)0733-9372(2005)131:9(1330).

Lawrence, M. G., T. M. Butler, J. Steinkamp, B. R. Gurjar, and J. Lelieveld. 2007. "Regional pollution potentials of megacities and other major population centers." *Atmos. Chem. Phys.* 7 (14): 3969–3987. https://doi.org/10.5194/acp-7-3969-2007.

Moonsammy, S., T. D. T. Oyedotum, O. Oderinde, M. Durojaiye, and A. Durojaye. 2021. "Exhaust determination and air-to-fuel ratio performance of end-of-life vehicles in a developing African country: A case study of Nigeria." *Transp. Res. D Transp. Environ.* 91 (Feb): 102705. https://doi.org/10.1016/j.trd.2021.102705.

Ogunseye, O. O., G. R. E. E. Ana, D. C. Uhiara, and D. G. Shendell. 2018. "Carboxyhaemoglobin levels among traders exposed to vehicular emissions in three motor parks in Ibadan, Nigeria." *J. Environ. Public Health* 3 (Jun): 1–8. https://doi.org/10.1155/2018/9174868.

OICA (International Organization of Motor Vehicle Manufacturers). 2019. "Global sales statistics of passenger cars." Accessed January 10, 2020. https://www.oica.net/category/sales-statistics/.

Pandey, A., G. Pandey, and R. K. Mishra. 2016. "Tailpipe emission from petrol driven passenger cars." *Transp. Res. D Transp. Environ.* 44 (May): 14–29. https://doi.org/10.1016/j.trd.2016.02.002.

Perappadan, B. S. 2012. "Air quality in Delhi is fast worsening." Accessed March 12, 2020. http://www.thehindu.com/todays-paper/tp-national/tp-newdelhi/air-qualityin-delhi-is-fast-worsening/article3499013.ece.

Rhys-Tyler, G. A., W. Legassick, and M. C. Bell. 2011. "The significance of vehicle emission standards for levels of exhaust pollution from light vehicles in an urban area." *Atmos. Environ.* 45 (19): 3286–3293. https://doi.org/10.1016/j.atmosenv.2011.03.035.

Rice, M. B., 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. 2015. "Long-term exposure to traffic emissions and fine particulate matter and lung function decline in the Framingham heart study." *Am. J. Respir. Crit. Care Med.* 191 (6): 656–664. https://doi.org/10.1164/rccm.201410-1875OC.

Sanchita, I., H. H. Masjuki, M. A. Kalam, I. M. Fattah-Rizwanul, M. M. Rashed, and H. K. Rashedul. 2014. "A review on idling reduction strategies to improve fuel economy and reduce exhaust emissions of transport vehicles." *Energy Convers. Manage.* 88 (Dec): 794–807. https://doi.org/10.1016/j.enconman.2014.09.036.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

Singh, A. K., H. K. Gupta, K. Gupta, P. Singh, V. B. Gupta, and R. C. Sharma. 2007. "A comparative study of air pollution in Indian cities." *Bull. Environ. Contam. Toxicol.* 78 (5): 411–416. https://doi.org/10.1007/s00128-007-9220-9.

Wang, H., L. Fu, Y. Zhou, X. Du, and W. Ge. 2010. "Trends in vehicular emissions in China's mega cities from 1995 to 2005." *Environ. Pollut.* 158 (2): 394–400. https://doi.org/10.1016/j.envpol.2009.09.002.

WHO (World Health Organization). 2019. "The global health observatory." Accessed January 7, 2020. https://www.who.int/data/gho/data/themes/theme-details/GHO/air-pollution.

Wong, Y. K., X. H. H. Huang, Y. Y. Cheng, P. K. K. Louie, A. L. C. Yu, A. W. Y. Tang, D. H. L. Chan, and J. Z. Yu. 2019. "Estimating contributions of vehicular emissions to PM2.5 in a roadside environment: A multiple approach study." *Sci. Total Environ.* 672 (Jul): 776–788. https://doi.org/10.1016/j.scitotenv.2019.03.463.

Downloaded from ascelibrary.org by Delhi Technological University on 05/27/22. Copyright ASCE. For personal use only; all rights reserved.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification

JRFID-0034-02-2022

Investigation of Circularly Polarized CPW fed Antenna as a 2.45 GHz RFID Reader

Amit Birwal, *Member, IEEE,* Vipul Kaushal, *Student Member, IEEE* and Kamlesh Patel, *Member, IEEE*

***Abstract*— In this paper, a bi-directional coplanar-waveguide (CPW)-fed antenna with circular polarization (CP) is investigated as a radio frequency identification (RFID) reader antenna. The antenna geometry is square and printed on one side of the FR4 substrate only, which consumes an overall dimension of 60×60 mm^2 with a thickness of 1.5 mm. The simulated and measured parameters of the antenna are in agreement with an impedance bandwidth (S_{11}<-10 dB) of 40.2% (2.18–3.28 GHz) and ARBW (<3 dB) of 28.3% (2.03-2.7 GHz), respectively. The peak gain of the proposed antenna is 3.17 dBi in both ± z directions. An Arduino UNO-based open-source microcontroller setup is used for studying this antenna as an RFID reader antenna, along with a wireless module NRF24L01 module at 2.45 GHz with an external antenna (tag). The path loss between the tag and reader antennas is found to increase in the desired band due to interference exists.**

***Index Terms*— Axial Ratio (AR), Circular Polarization (CP), RFID, Reader Antenna, and Return Loss.**

I. Introduction

The radio frequency identification (RFID) is commonly used in many technologies, with applications ranging from defense, healthcare, and many supply chain organizations [1]. RFID provides several advantages over traditional forms of identification, including higher detection accuracy, speed, and its unique feature to adapt to the surroundings. RFID systems are classified in low, high, and ultra-high frequency (LF, HF, and UHF) and microwave frequencies based on their frequency bands of operation. RFID has advanced considerably in recent years, particularly in the HF and UHF bands. Meanwhile, as smartphones and iPads become more popular, the use of wireless fidelity (Wi-Fi, interior communication) is increased. So, investigating RFID antennas in these domains is of a prime importance.

Usually, the antenna requirement for Microwave based RFID systems is characterized by wide impedance bandwidth, high gain, circular polarization (CP), etc. [2]. These characteristics are studied by many researchers in various types of microstrip antenna based on ceramic materials, dipoles, spiral, and planar inverted F antenna to be used as the reader of RFID systems [3-6]. Apart from such designs, circularly polarized antennas are also becoming popular in the Internet of Things (IoT) based systems [7]. The UHF, 920-925 MHz, and 2.45 GHz RFID reader antennas for IoT applications are reported in [8]. The RFID tag's reader antenna is an integral component of the RFID system. Due to multipath signal reception, fading occurs, resulting in signal power loss and thus impacting the system reliability. Using CP-based microstrip antennas, such fading impact between the reader antenna and the tag antenna can be eliminated, and many useful applications can be designed by using frequency bands reserved for Industrial, Scientific, and Medical (ISM) applications. To integrate the reader antenna with other IoT applications such as Wi-Fi and LTE, a wideband CP antenna is required, which can support a wide bandwidth from 2 to 2.5 GHz as shown in Fig. 1.

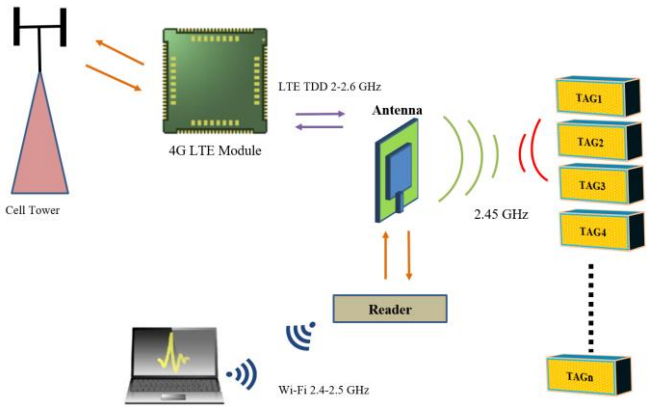


Fig. 1. Antenna with IoT system architecture.

A coplanar-waveguide (CPW)-fed antenna with circular polarization (CP) has been reported earlier to cover a wide CP bandwidth [9], whereby incorporating an inverted L-shaped arm at the left side and a tuning stub at the right side of the ground plane in a conventional CPW-fed monopole antenna, a wide impedance bandwidth and axial-ratio bandwidth (ARBW) is achieved. A symmetric-aperture antenna proposed in [10] uses a CPW-fed and a pair of the diamond-shaped aperture is used to generate a CP radiation in both ±z directions. In this extended work, various parameters such as input impedance, return loss (S_{11}), axial ratio, and gain are measured. This antenna is integrated with an Arduino UNO-based open-source microcontroller setup is proposed along with a wireless module

This work was supported in part by the Faculty Research Programme (FRP) of the Institute of Eminence (IoE) scheme of the University of Delhi under Grant ref. No. IoE/2021/12/FRP dated 29/10/2021.

Corresponding author: K. Patel (kpatel@south.du.ac.in)

A. Birwal is with Department of Electronic Science, University of Delhi South Campus, New Delhi-110021, India.
(e-mail: amit.birwal@south.du.ac.in).

V. Kaushal is with Department of Electronic Science, University of Delhi South Campus, New Delhi-110021, India
(e-mail: vipul.kaushal24@gmail.com).

K.Patel is with Department of Electronic Science, University of Delhi South Campus, New Delhi-110021, India (e-mail: kpatel@south.du.ac.in).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification

NRF24L01 module in conjunction with another tag antenna, and found to be suitable for RFID reader antenna where bi-directional coverage is required. The proposed antenna provides a right-hand circular polarization (RHCP) radiation in boresight (+z) and left-hand circular polarization (LHCP) in backside (-z) directions.

II. ANTENNA DESIGN

In this section, the evolution stages of the antenna are discussed.

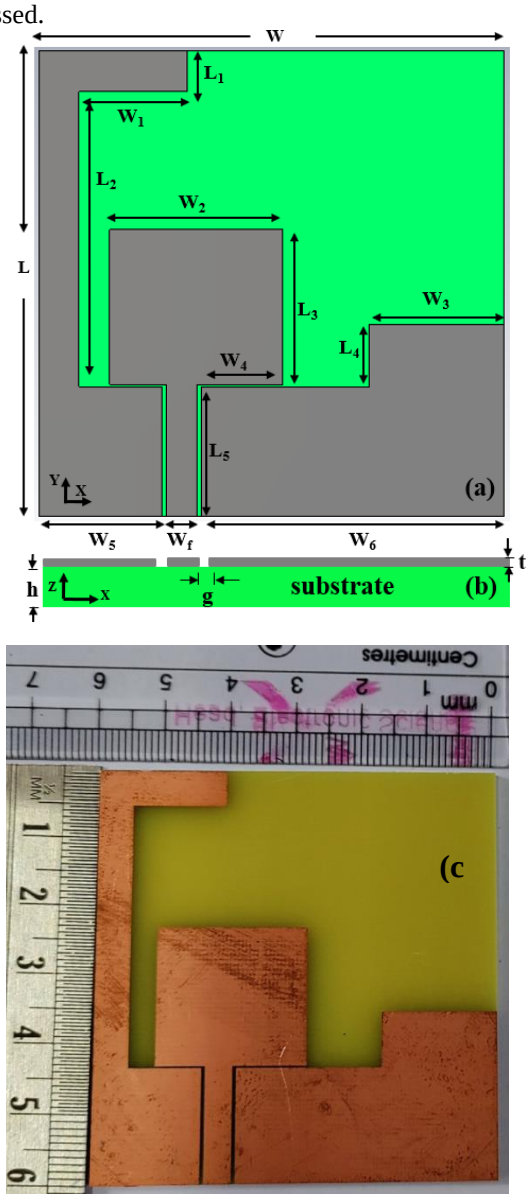
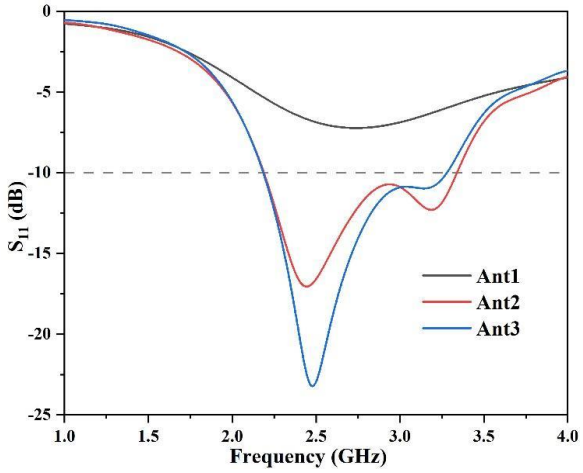


Fig. 2. Geometrical view of the proposed antenna (a) top layer view (b) side view (c) fabricated prototype.

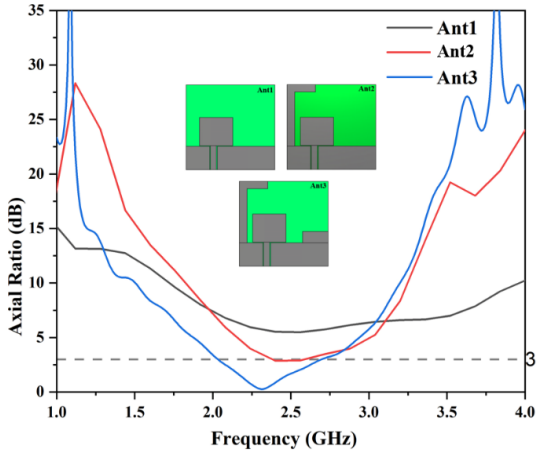
In the first stage, a conventional CPW-fed monopole antenna (Ant1) is designed. To improve the impedance bandwidth and shift the resonance towards the lower ISM band i.e., 2.45 GHz, an inverted L-shaped stub is introduced in the ground plane making a defected ground antenna (Ant2). To incorporate the circular polarization, an additional tuning stub is added at the right side of the ground plane (Ant3). The S_{11} (dB) and axial ratio (dB) during these evolution stages are shown in Fig. 3. The final optimized geometry of the proposed antenna is shown in Fig. 2, which shows the top and side views of the antenna. Table 1 lists all the optimized dimensions of this antenna, it can be seen that the overall geometry of this antenna is 60×60×1.5 mm^3. To obtain the wideband circular polarization (CP) in the desired band, the dimensions W_3×L_4 of the rectangular tuning stub are optimized.

Table 1. Optimized dimensions of the CP antenna.

Parameter	Value (mm)	Parameter	Value (mm)
W	60	W_1	14
L	60	W_2	22.3
L_1	5.2	W_3	17.4
L_2	31.8	W_4	11
L_3	20	W_5	15.9
L_4	8	W_6	38.9
L_5	16.7	W_f	1.5
g	0.57	t	0.035



(a)



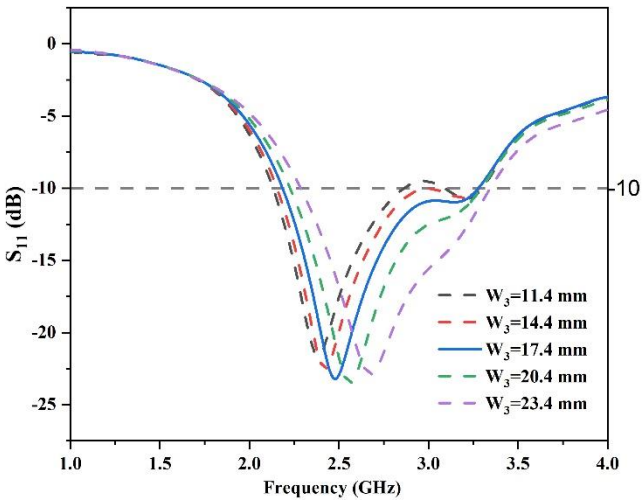
(b)

Fig. 3. Antenna characteristics during evolution stages (a) S_{11} (dB) (b) Axial ratio (dB).

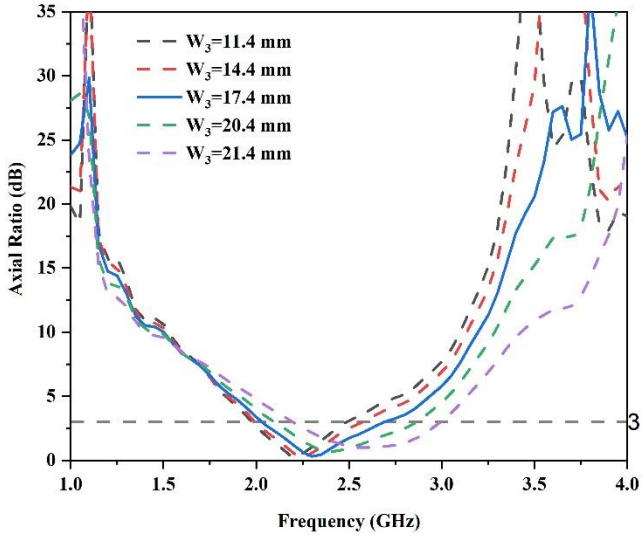
A parametric analysis of tuning stub length W_3 is carried out to achieve a wideband CP as shown in Fig. 4. Here the length W_3 is varied from 11.4 mm to 23.4 mm in steps of 3 mm.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification



(a)



(b)

Fig. 4. Parametric study of parameter W_3 on antenna parameters (a) S_{11}(dB) and (b) Axial ratio (dB).

For an initial W_3 value of 11.4 mm, narrow ARBW is achieved from 1.97- 2.49 GHz i.e. 0.52 GHz at a resonance of 2.2 GHz. On increasing the length of the tuning stub to 14.3 mm, the resonance shifts to 2.24 GHz with an ARBW of 0.59 GHz. A wide ARBW of 0.7 GHz with a peak axial ratio of 0.2 dB and impedance bandwidth of 2.18–3.28 GHz at a resonance of 2.45 GHz is observed for W_3 of 17.4 mm, which is within the ARBW range. On further increasing W_3 to 20.4 to 23.4 mm, there is a minor increase in ARBW, but the peak axial ratio value drops to 0.75 dB, and the resonance at the impedance bandwidth shifts to 2.57 GHz. Considering the design requirement of microwave RFID at 2.45 GHz, the final optimized value of W_3=17.4 mm is selected. This optimized value is helping the antenna to generate two orthogonal magnetic current components with equal magnitude and 90° phase difference.

The substrate used is FR4 (dielectric constant ε_r=4.5, height h=1.5 mm, and loss tangent $tan\,\delta = 0.02$). The central width (W_f) and gap width (g) of the CPW feed line are optimized to provide perfect matching at the resonance frequency. Fig. 5 shows the surface current distribution on the proposed antenna when port one is excited at 2.45 GHz. When a phase zero signal is applied, the black color dominant current vector can be seen along the +x direction and changes its orientation to +y when the phase changes to 90^0. Similarly, the orientation changes to -x and -y for phases 180^0 and 270^0, respectively. If we see the rotation of this resultant current vector from the antenna top i.e., +z the sense of rotation is counterclockwise, which is equivalent to right-hand circular polarization (RHCP).

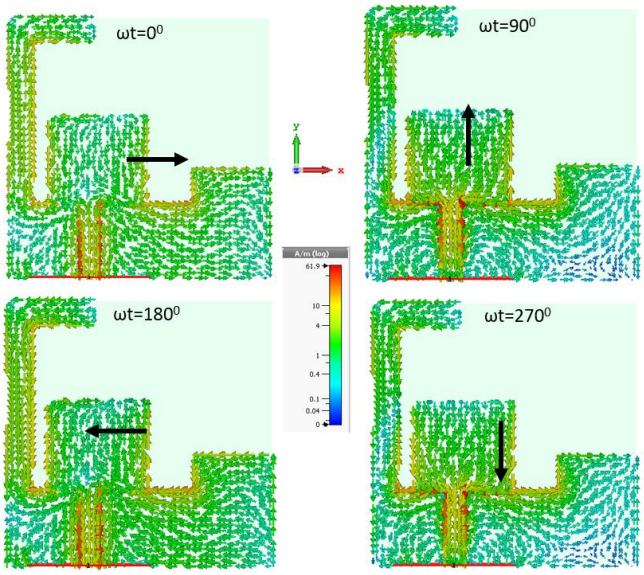


Fig. 5. Surface current distribution on the proposed antenna at 2.45 GHz at four different time instants.

To better understand the CP mechanism offered by this antenna the electric field distribution available on the antenna aperture is plotted using the commercial CST Microwave studio[11], where a time-varying electric field pattern is plotted at different phase instants i.e., $\omega t = 0^0$, $90^0, 180^0, 270^0$ as shown in Fig. 6. At $\omega t = 0^0$, the electric field dominant vector is along the +y direction, and its moves to the -x direction at a progressive phase of $\omega t = 90^0$. Similarly, the dominant electric field vector goes from -y to +x during phase instants $\omega t = 180^0$ and $\omega t = 270^0$, respectively. As a result, it shows that the field vector resembles the counterclockwise direction at 2.45 GHz, which corresponds to RHCP radiation in the +z direction.

Fig. 7 shows the input impedance (Z_{in}) variation with frequency, and it can be observed that this antenna provides an optimal matching within the band (2.18-3.28 GHz), where the real impedance is almost near 50 Ω and the imaginary part is less than 25 Ω to almost near zero.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification

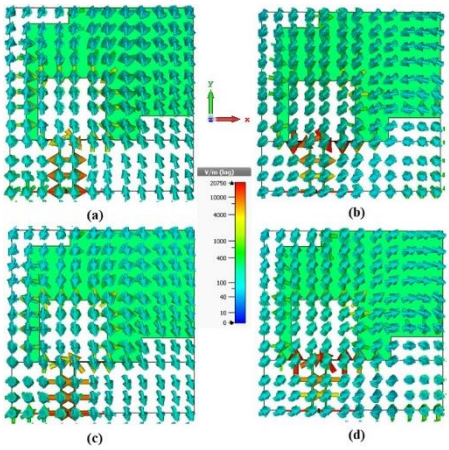


Fig. 6. Electric field distribution on the aperture of the antenna at 2.45 GHz for different phases (a) $\omega t = 0^0$(b) $\omega t = 90^0$(c) $\omega t = 180^0$(d) $\omega t = 270^0$

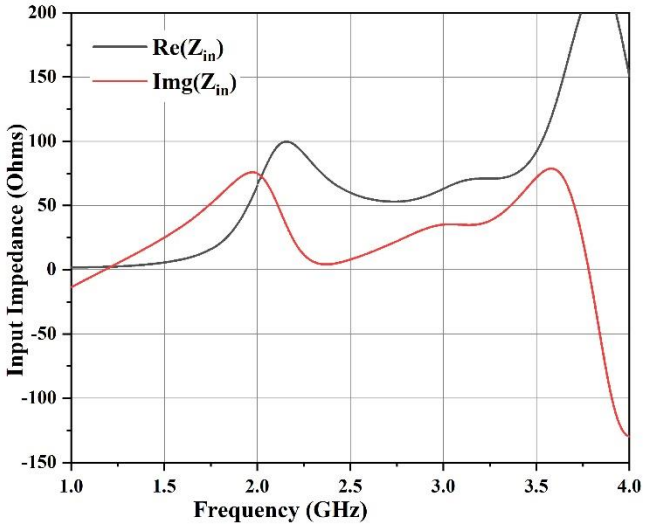
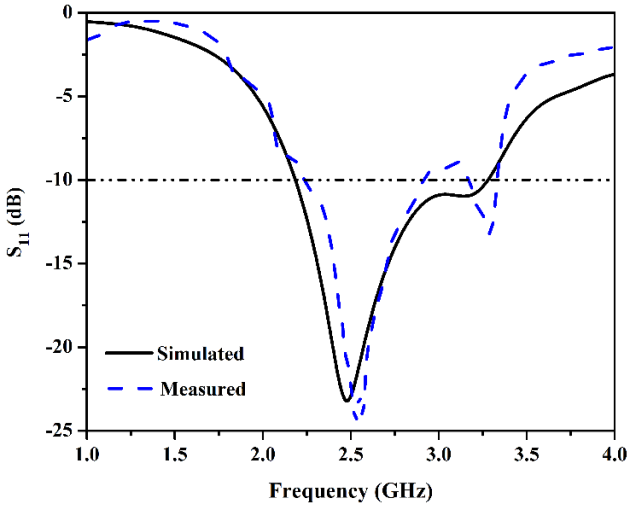


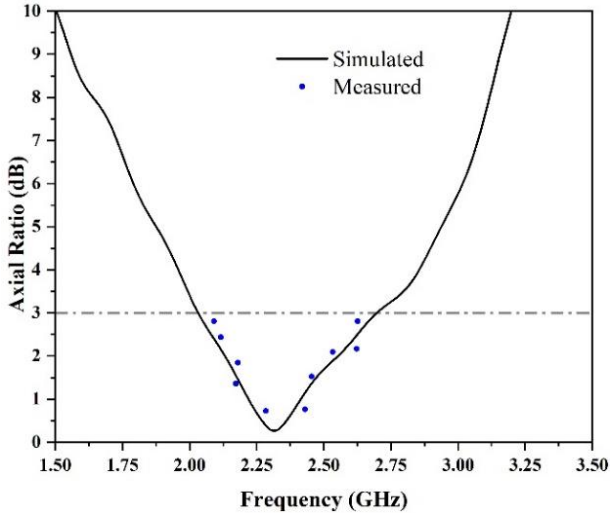
Fig. 7. Simulated input impedance (Z_{in}) vs frequency showing real and imaginary parts.

III. RESULTS AND DISCUSSION

Some geometrical parameters have been optimized to study the performance of this antenna in the CST studio suite and fabricated using an LPKF ProtoMat S100 PCB milling machine. To carry out the measurement, a handheld vector network analyzer from Rohde & Schwarz ZVH8 is used. The (S_{11}<-10 dB) of the antenna is measured and plotted along with the simulated results in Fig. 8 with an impedance bandwidth of around 1.1 GHz with a resonance peak at 2.5 GHz and 3.2 GHz. The second peak of about -12 dB is found due to the second resonance at 3.2 GHz which might have currents in the same direction as for the first resonance at 2.5 GHz, with a different current distribution. This shows that the capability of this antenna is suitable for RFID application in the ISM frequency band i.e., 2.45 GHz. The 3-dB axial ratio bandwidth (ARBW) is also measured at a few points near 2.45 GHz and found in close agreement with the simulated as shown in Fig. 8(b).



(a)



(b)

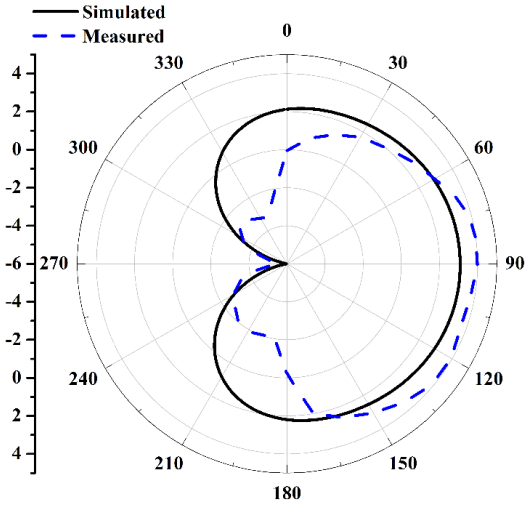
Fig. 8. Simulated and measured result (a) S_{11} (dB) (b) axial ratio (dB) of the proposed antenna.

The ARBW is falling within the range of impedance bandwidth that confirms that the antenna is circularly polarized within the operational bandwidth and is a good candidate to be used as an RFID reader antenna.

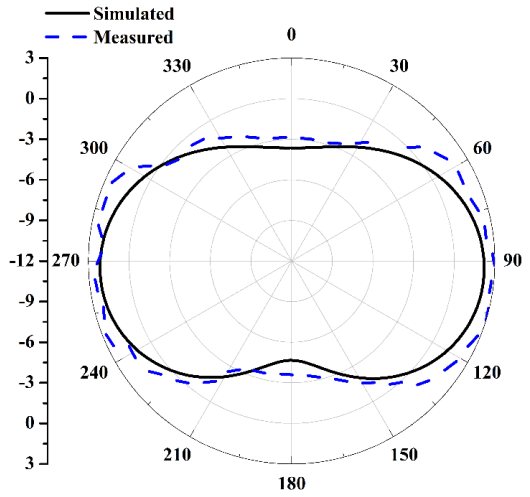
Fig. 9 shows the stable radiation pattern measurement results in both XY and YZ planes. Table 2 show a comparison of some reported CPW-fed CP antenna published in the literature, where the measured data is collected from these reported antennas. It is observed that the proposed antenna is compact and has comparable impedance bandwidth and ARBW with other reported designs. Also, it has been observed that the gain of the antenna is around 3 dBi at 2.45 GHz and an increase up to 4 dBi at 3.1 GHz is observed, as shown in Fig. 10.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification



(a)



(b)

Fig. 9. 2D radiation pattern of the antenna at 2.45 GHz (a) XZ-plane, (b) YZ plane

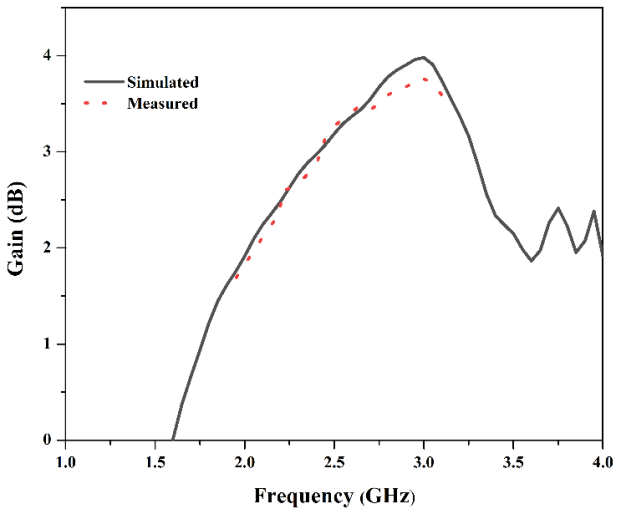


Fig. 10. Measured and simulated gain at boresight.

Table 2. Comparison of proposed CP antenna with other reported antennas in literature.

Ref. Antenna	Size (mm^2)	3-dB ARBW (%)	10-dB return loss bandwidth (%)	Gain (dBi)
[3]	120 × 120	47	64	<3.4
[16]	70 × 70	12	39	<2.5
[17]	114 × 110	24	35	<3.6
[18]	112 × 112	51	56	<3.8
[19]	250 × 250	16	22	<9.8
[20]	80 × 80	22	34	<5.8
This work	60 × 60	30	42	<3

IV. ANTENNA INTEGRATION

To test the capability of the proposed antenna to be used as a reader antenna, an Arduino UNO-based open-source microcontroller [12] setup is proposed along with a wireless module NRF24L01 module [13] at 2.45 GHz with an external antenna. The microcontroller board consists of multiple input-output pins for interfacing with external circuits, which is NRF24L01 in our case. Serial Peripheral Interface (SPI) is used between NRF24L01 and Arduino to realize a wireless communication system. This wireless module NRF24L01 can support multiple output power values (0 dBm, -6 dBm, -12 dBm, or -18 dBm) radiated by an external antenna at 2.45 GHz. In the setup shown in Fig. 11, we used two setups, one acting as a transmitter and the other acting as a receiver.

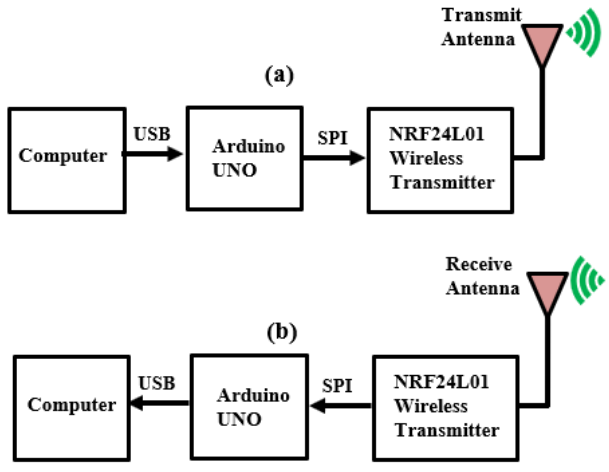


Fig. 11. Antenna integration setup (a) transmit path and (b) receive path.

To realize an RFID system shown in Fig. 1, transmit antenna of NRF24L01 is replaced with the proposed CPW antenna, and the folded microstrip antenna is used as a receive antenna reported in [14], which resembles a tag antenna of size

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification

49×39 mm^2 and resonates at 2.45 GHz as shown in Fig. 12. The reported gain of the antenna is around 2.5 dBi and it is termed a tag antenna or simple tag.

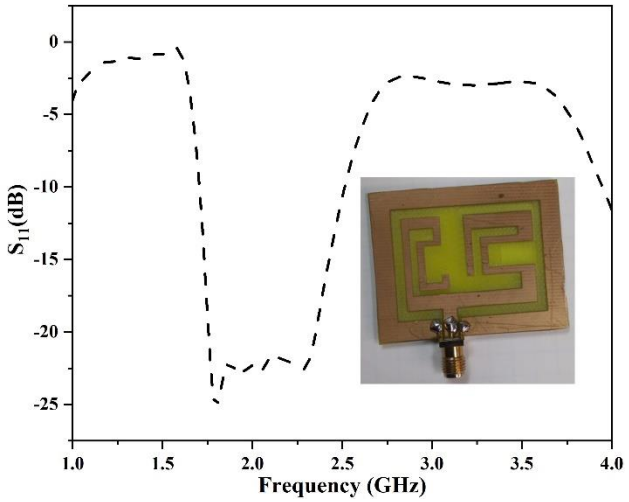


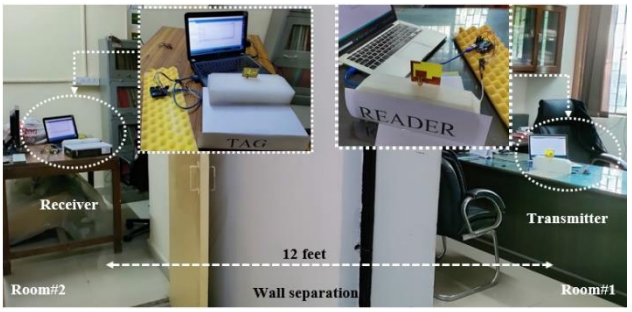
Fig. 12. Measured return loss S_{11} (dB) of the tag antenna.

To integrate and configure the wireless communication system as shown in Fig 13(a), a C program is written and uploaded in Arduino. The program is written with a dummy text message which is broadcast by the reader antenna at a defined transmit power level Pt. The receiver or tag is placed in another room at a distance of 12 feet, which received the message transmitted from the reader. Here the antennas are configured in such a way that the reader will only operate in transmit mode and the receiver will only operate in receive mode. The same setup is reproduced in the anechoic chamber to minimize the unwanted reflection sources, as shown in Fig. 13(b). It has been also observed that the tag is receiving the dummy message transmitted from the reader at each angular location when the tag antenna is varied from 0^0 to 360^0 in a step of 45^0, this is due to the bi-directional radiation of the CPW antenna.

Table 3. Power received for varying transmit power and distance.

Distance (cm) D	Pr @ Pt = -18dBm	Pr @ Pt = -12dBm	Pr @ Pt = -06dBm
50	-32.62	-30.2	-27.38
100	-39.92	-34.39	-32.05
150	-44.02	-37.98	-34.63
200	-45.93	-40.76	-38.46

The distance between the reader and the tag antenna (d) is varied from 50 to 200 cm in a step of 50 cm. The transmit power Pt is taken as a variable in the C program for three values, -18 dBm, -12 dBm, and -6 dBm and the corresponding received power is measured using Rohde & Schwarz FSL6 spectrum analyzer after disconnecting the tag antenna from the NRF24L01 module. All the corresponding received power is measured and presented in Table 3. It is observed from Fig. 14 that by increasing the transmit power Pt, the distance between the reader and tag (i.e., read range) can be increased.



(a)



(b)

Fig. 13. Proposed antenna integration with Arduino (a) external room setup (b) anechoic chamber setup.

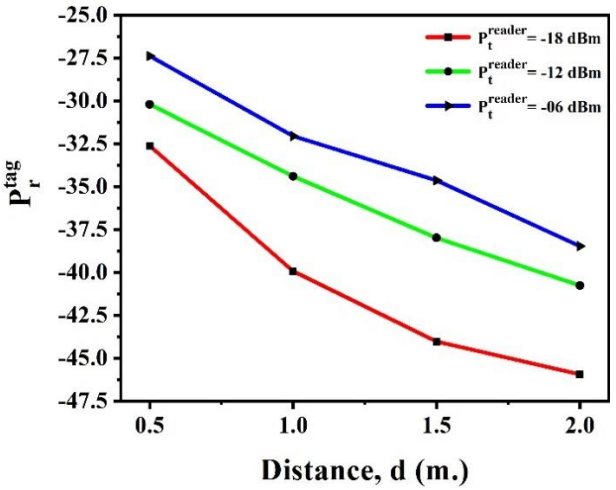


Fig. 14. Received power (dBm) for varying distances.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification

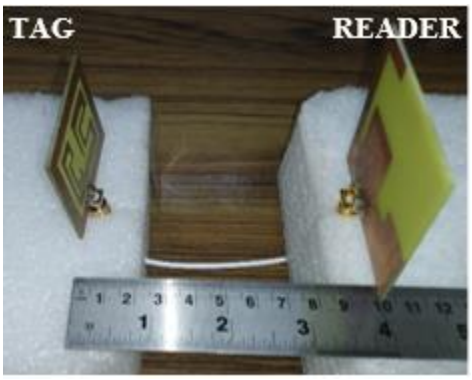


Fig. 15. Path loss measurement setup between tag and reader antennas in the lab environment.

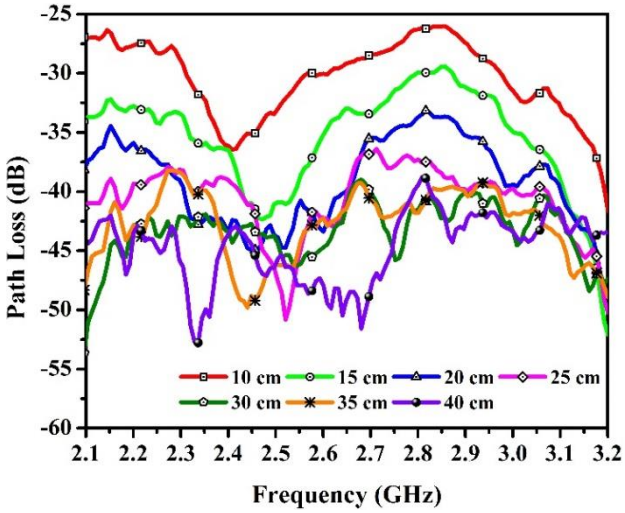


Fig. 16. Measured Path loss between tag and reader antennas for varying distances.

The path loss (PL) between the tag and reader antennas is obtained from S-parameters measurements by varying the distance between them from 10 cm to 40 cm in a 10 cm step [15]. The PL values measured in the operating frequency range of the reader antenna i.e. 2.1 GHz to 3.2 GHz are given in Fig. 16 and found to be increasing as the distance between tag and reader antennas increases. In addition, a rise in PL is observed in the 2.45 GHz band which confirms the interference with other applications running on 2.45 GHz i.e. Wi-Fi and Bluetooth in the lab environment. So, enhancing the reader power is a possible solution to reduce the interference or incorporate the EMC solutions in such an environment.

V. Conclusion

This paper presents a square shape CPW-fed antenna providing a bidirectional radiation pattern. The antenna is circularly polarized and suitable for RFID reader applications in the ISM band at 2.45 GHz. The geometry of the antenna is simple and printed on only one side of the low-cost FR4 substrate with an overall dimension of 60×60×1.5 mm^3. The measured results confirmed an impedance bandwidth (S_{11}<-10 dB) from (2.1–3.2 GHz) and ARBW (<3 dB) from (2-2.7 GHz), respectively with a gain of 3 dBi at 2.5 GHz. The antenna is also tested as a reader antenna in transmit mode using a wireless communication system, where multiple scenarios are tested with a tag antenna. The antenna can also be useful for many other IoT-based applications where bidirectional RFID coverage and integration with other technologies like Wi-Fi, LTE, etc., is required.

References

[1] Gao, Yanchen, D. Yang, and W. Ning, "RFID application in tire manufacturing logistics," in *2010 IEEE International Conference on Advanced Management Science (ICAMS 2010)*, vol. 3, Chengdu, 2010, pp. 109-112.

[2] Son, Wang-Ik, H. L. Lee, M. Q. Lee, S. B. Min, and J. W. Yu, "Compact square quadrifilar spiral antenna with circular polarization for UHF mobile RFID reader," In *2010 Asia-Pacific Microwave Conference, IEEE*, Yokohama, Japan, 2011, pp. 2271-2274.

[3] R. Cao, and S. C. Yu, "Wideband compact CPW-fed circularly polarized antenna for universal UHF RFID reader," *IEEE Transactions on Antennas and Propagation*, vol. 63, no. 9, pp. 4148-4151, Jun. 2015, doi: 10.1109/TAP.2015.2443156.

[4] T. N. Chang, and J. M. Lin, "A novel circularly polarized patch antenna with a serial multislot type of loading," *IEEE transactions on antennas and propagation*, vol. 55, no. 11, pp. 3345-3348, Nov. 2007, doi: 10.1109/TAP.2007.908846.

[5] W. G. Lim, W.I. Son, K. S. Oh, W. K. Kim, and J. W. Yu, "Compact integrated antenna with circulator for UHF RFID system," *IEEE Antennas and Wireless Propagation Letters*, vol. 7, pp. 673-675, Mar. 2008, doi: 10.1109/LAWP.2008.921348.

[6] R. C. Hua, and T.G. Ma, "A printed dipole antenna for ultra high frequency (UHF) radio frequency identification (RFID) handheld reader," *IEEE Transactions on Antennas and Propagation*, vol. 55, no. 12, pp. 3742-3745, Dec. 2007, doi: 10.1109/TAP.2007.910521.

[7] A. Birwal, S. Singh, B. K. Kanaujia and S. Kumar, "Broadband CPW-fed circularly polarized antenna for IoT-based navigation system," *International Journal of Microwave and Wireless Technologies*, vol. 11, no. 8, pp. 835-843, May 2019, doi: https://doi.org/10.1017/S1759078719000461.

[8] A. Romputtal and C. Phongcharoenpanich, "IoT-linked integrated NFC and dual band UHF/2.45 GHz RFID reader antenna scheme," *IEEE Access*, vol. 7, pp. 177832-177843, Dec. 2019, doi: 10.1109/ACCESS.2019.2958257.

[9] A. Birwal, V. Kaushal, and K. Patel, "Circularly Polarized Broadband Co-Planar Waveguide fed Antenna for 2.45 GHz RFID Reader," in *2021 IEEE International Conference on RFID Technology and Applications (RFID-TA)*, Delhi, India, 2021, pp. 109-112.

[10] Nasimuddin, Z. N. Chen, and X. Qing, "Symmetric-Aperture Antenna for Broadband Circular Polarization," *IEEE Transactions on Antennas and Propagation*, vol. 59, no. 10, pp. 3932-3936, Oct. 2011, doi: 10.1109/TAP.2011.2163757.

[11] *"CST Microwave Studio, Version 2017," 2017.*

[12] "*Arduino Uno Rev3*," [Online]. Available: https://store-usa.arduino.cc/products/arduino-uno-rev3/.

[13] "*nRF24 Series*," [Online]. Available: https://www.nordicsemi.com/Products/nRF24-series.

[14] A. Birwal, S. Singh and B. K. Kanaujia, "Smart compact-folded microstrip antenna for GSM, LTE, and WLAN applications," In Smart Systems and IoT: Innovations in Computing (pp. 475-481). Springer, Singapore.

[15] V. Kaushal, A. Birwal, S. M. Patel, and K. Patel, "Path Loss of Two-Port Circular-Ring Slot Antenna For RFID Applications," in *2021 IEEE International Conference on RFID Technology and Applications (RFID-TA)*, Delhi, India, 2021, pp. 120-123.

[16] C. C. Chou, K. H. Lin, and H. L. Su, "Broadband circularly polarized cross-patch-loaded square slot antenna," *Electron. Lett.*, vol. 43, no. 9,pp. 485–486, Apr. 2007, doi: 10.1049/el:20070189.

[17] S.Zhang, H. Huang, and Y. Yin, "A broadband CPW-fed circularly polarized square slot antenna for UHF RFID applications," *Progress In Electromagnetics Research C*, vol. 50, pp. 39-46, May 2014, doi: 10.2528/PIERC14021707.

[18] W. Abdelrahim, and Q. Feng, "Compact broadband dual-band circularly polarised antenna for universal UHF RFID handheld reader and GPS applications," *IET Microw. Antennas Propag.*, vol. 13, pp. 1664-1670, Aug. 2019, doi: 10.1049/iet-map.2018.5970.

[19] Liu, XiongYing, Yi Liu, and Manos M. Tentzeris, "A novel circularly polarized antenna with coin-shaped patches and a ring-shaped strip for worldwide UHF RFID applications," *IEEE Antennas Wirel. Propag. Lett.*, vol. 14, pp. 707–710, Dec 2014, doi: 10.1109/LAWP.2014.2378513.

[20] C.W. Chiu, C. A. Ou, and H. C. Wang, "Compact printed quadrifilar helix antenna for universal RFID hand-held reader," *J. Electromagn. Waves Appl.*, vol. 29, pp. 891–904, Sep. 2014, doi: 10.1080/09205071.2015.1026407.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JRFID.2022.3172691, IEEE Journal of Radio Frequency Identification



A. Birwal (Member, IEEE) received his B. Tech. degree in Electronics and Communication Engineering from Guru Gobind Singh Indraprastha University, Delhi, India in 2003 and M. Tech. degree in Microwave Electronics from the University of Delhi, India in 2006. He is a Ph.D. from the University of Delhi, New Delhi, India, in 2021. He worked at various levels in the development and test phases of telecom devices at Ericsson India. He now holds the position of Assistant Professor at the Department of Electronic Science, University of Delhi South Campus, India. His main research interests are designing and optimizing planar antennas for IoT-based systems, communications systems, etc.



V. Kaushal received his B.Sc. degree in Electronics and M.Sc. degree in Electronics from the University of Delhi in 2017 and 2019, respectively. Presently, he is a research scholar with the Department of Electronic Science, University of Delhi South Campus, New Delhi. His research interests are MIMO antenna designs and their analysis for 5G communications.



K. Patel (Member, IEEE) received his M.Sc. degree in Electronics and M.Tech. degree in Microwave Electronics from Rani Durgavati Vishwavidyalaya, Jabalpur, and University of Delhi, India in 1999 & 2003, respectively. He is Ph.D. from Delhi Technological University, Delhi. He was a Scientist with CSIR- National Physical Laboratory from 2004 to 2013. Since 2013, he is an Assistant Professor in the Department of Electronic Science, University of Delhi South Campus, New Delhi. His research interests include microwave components, material characterization, planar antennas for mobile communications, etc.

2469-7281 (c) 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.
Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 09:42:33 UTC from IEEE Xplore. Restrictions apply.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Investigation of electronic, mechanical and thermoelectric properties of quaternary Heusler compounds ZrRhTiZ (Z = In, Al)

Vivek Kumar, Mohit Kumar, Mukhtiyar Singh *

Department of Applied Physics Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Heusler alloys
Wien2k
Half-Metallicity
Mechanical properties
Electronic properties, Thermoelectric

ABSTRACT

In the present work, a first-principles investigation of the electronic, mechanical and transport properties of quaternary Heusler compounds ZrRhTiZ (Z = In, Al) is reported. The observed lattice constants for ZrRhTiIn and ZrRhTiAl are 6.62Å and 6.44Å, respectively, which are in good agreement with earlier work. The alloys are semiconducting in spin-down state with indirect bandgap 0.39 eV and 0.30 eV for ZrRhTiIn & ZrRhTiAl, respectively but metallic in spin-up state which make them half-metallic ferromagnets. Various thermoelectric properties namely, Seebeck coefficient, electrical and thermal conductivities, and power factor are reported in the paper. The highest observed values of power factors are $1.57x10^{11}$ $Wm^{-1}K^{-2}s^{-1}$ and $\sim 0.66x10^{11}$ $Wm^{-1}K^{-2}s^{-1}$ for ZrRhTiAl and ZrRhTiIn alloys, respectively. All these properties indicate that these materials can be employed in spintronic and thermoelectric devices.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

1. Introduction

Heusler alloys (HA) have recently attracted scientific attention due to their unique structural, electrical, magnetic, mechanical, and thermoelectric properties [1–3]. Among the family of Heusler alloy, Equiatomic Quaternary Heusler alloy (EQH) is one of the important categories due its application in the spintronics, magnetoelectric, magnetocaloric, ferromagnetic shape memory alloys[4–6]. EQH alloys crystallise with the Y-structure, which are more suitable for various applications than other HA because of low power dissipation. HA can be classified into three categories depending upon the stoichiometry, viz. (1) Half-Heusler (XYZ), (2) full-Heusler (X_2YZ) and (3) quaternary Heusler (XX'YZ), where X, X', and Y are transition metal elements, and Z is an element from the main group [7,8]. The EQH has space group 216-*F43m* with atoms X and X' at the 4a (0, 0, 0) and 4b (0.5, 0.5, 0.5) positions, atom Y at the 4c (0.25, 0.25, 0.25) position, and main group element Z at the 4d position (0.75, 0.75, 0.75) [9]. The occupation rule aids in the formation of a highly ordered rather than disorderly structure. Various quaternary Heusler compounds have been described as novel Half-metals using both experimental and theoretical studies. The ZrFeTiZ (Z = Ge, Al, Si), ZrFeTiZ (Z = In, Al, Ga), and ZrCoTiZ (Z = Ga, Al, Ge, Si) are Zr-based EQH alloys that have been explored for spin-gapless semiconductors [10,11].

Heusler alloy has thermoelectric properties that allows it to transform heat energy into electrical energy (Seebeck effects) [12]. It's a promising attribute of HA since it can be exploited in electronic refrigeration and power generating applications that transfer heat to electricity[13,14]. Certain parameters, such as the Seebeck coefficient, electrical and thermal conductivities as a function of chemical potential and temperature, can be investigated to estimate the thermoelectric efficiency. This study uses DFT and the BoltzTraP code to report on the structural, electrical, mechanical, and thermoelectric properties of ZrRhTiZ (Z = In, Al).

2. Computational methodology

The Density functional theory (DFT) was used to determine various parameters of the ZrRhTiZ (Z = In, Al) alloys. In the work, the extremely accurate all electrons linearized augmented plane wave approach developed in the WIEN2k code was used. The Perdew–Burke–Ernzerhof generalised gradient approximation (GGA-PBE) and Tran–Blaha modified Becke–Johnson (TB-mBJ) methods were used to derive the results. When the GGA approach has been

* Corresponding author.
E-mail addresses: mukhtiyarsingh@dtu.ac.in, msphysik09@gmail.com (M. Singh).

https://doi.org/10.1016/j.matpr.2022.04.483
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

Please cite this article as: V. Kumar, M. Kumar and M. Singh, Investigation of electronic, mechanical and thermoelectric properties of quaternary Heusler compounds ZrRhTiZ (Z = In, Al), Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.483

employed, the band gap results have been underestimated. To overcome erroneous estimated values, the Tran-Blaha modified Becke–Johnson approach was used. The structural relaxation for different compositional occupancy has been achieved with RK_{max} (plane wave cutoff) is set to 7 and a converged ground state is created utilising a fine k-mesh with 8000 k-point in the Full Brillouin-zone. Finer convergence has been achieved for the energy and charge estimation at the cost of the convergence threshold 0.0001 Ry and 0.001 e, respectively. In addition, the separation of valence and core states has a cut-off energy of −6.0 Ry. Highest vector magnitude in charge density Fourier expansion (G_{max}) = 13. The Boltzmann transport theory (BoltzTraP code) was used to determine the transport parameters of EQH ZrRhTiZ (Z = In, Al).

3. Results and discussion

3.1. Structural properties

The quaternary Heusler alloy XX'YZ are reported to have 3 possible configurations depending on how the atoms are arranged. The positions of X, X', Y, and Z atom are at 4c, 4d, 4b, and 4a; 4b, 4d, 4a and 4c; 4c, 4b, 4d and 4a, respectively [15]. The current ZrRhTiZ (Z = In, Al) is found to be most stable in Type-1. . The energy minimization approach was used to find the equilibrium lattice parameter. The Birch-Murnaghan equation is explored to determine the lattice parameter and the energy for ground state at equilibrium volume. The obtained equilibrium lattice parameters are found to be 6.63 Å and 6.45 Å, respectively for ZrRhTiIn and ZrRhTiAl. Fig. 1(a) and (b) shows the volume vs energy curve of ZrRhTiIn and ZrRhTiAl respectively.

3.2. Electron density

Understanding the chemical characteristics of elements and the bonding between atoms relies heavily on electron density charts. Fig. 2 (a) and (b) shows the electron density contour plot of both the EQH's along the 110 Plane, which helps us to know the chemical bonding in the alloys. Electron density plots also help us to understand the d-d hybridization via contour plots. It can be observed that the charge accumulations between the Zr-Ti atoms is comparatively less than Rh-Ti atoms and Rh-Al/In atoms indicating in Zr-Ti interacting via ionic bond while Rh-Ti and Rh-Al/In interacting via covalent band.

3.3. Mechanical properties

$c_{11} - c_{12} > 0, c_{11} + 2c_{12} > 0, c_{44} > 0$ Within the born-stability conditions, the elastic parameters were estimated to forecast the stability of the ZrRhTiZ (Z = In, Al) alloy [16]. i.e., The estimated elastic constants are shown in Table 1. It is observed that each material satisfies the stability conditions hence are mechanically stable. The following formula is used to compute the bulk modulus B [17];

$$B = \frac{C_{11} + C_{12}}{3} \tag{1}$$

The elastic constants are used to compute the shear modulus (G), Young's modulus (E), and compressibility, which are calculated using the conventional relationships listed below [17].

$$E = \frac{9BG_H}{3B + G_H} \tag{2}$$

$$G = C_{44} \tag{3}$$

$$\beta = \frac{1}{B} \tag{4}$$

The Poisson ratio (ν) is a measurement of a material's ductility or fragility, and is represented as follows [17]:

$$\nu = \frac{1}{2} - \frac{E}{6B} \tag{5}$$

Furthermore, the Zener's value (A) can be calculated using the following equation [17]:

$$A = \frac{2c_{44}}{c_{11} - c_{12}} \tag{6}$$

It is commonly understood that if A equals one unit, the system is isotropic and anisotropic otherwise.

Reuss shear modulus, Cauchy's pressure, Voigt shear modulus, and Hill shear modulus all have the same formula [17].

$$\boldsymbol{G_p} = \boldsymbol{C_{12}} - \boldsymbol{C_{44}} \tag{7}$$

$$\boldsymbol{G_R} = \frac{5(\boldsymbol{C_{11}} + \boldsymbol{C_{12}})\boldsymbol{C_{44}}}{4\boldsymbol{C_{44}} + 3(\boldsymbol{C_{11}} - \boldsymbol{C_{12}})} \tag{8}$$

$$\boldsymbol{G_v} = \frac{\boldsymbol{C_{11}} - \boldsymbol{C_{12}} + 3\boldsymbol{C_{44}}}{5} \tag{9}$$

$$\boldsymbol{G_H} = \frac{\boldsymbol{G_R} + \boldsymbol{G_v}}{2} \tag{10}$$

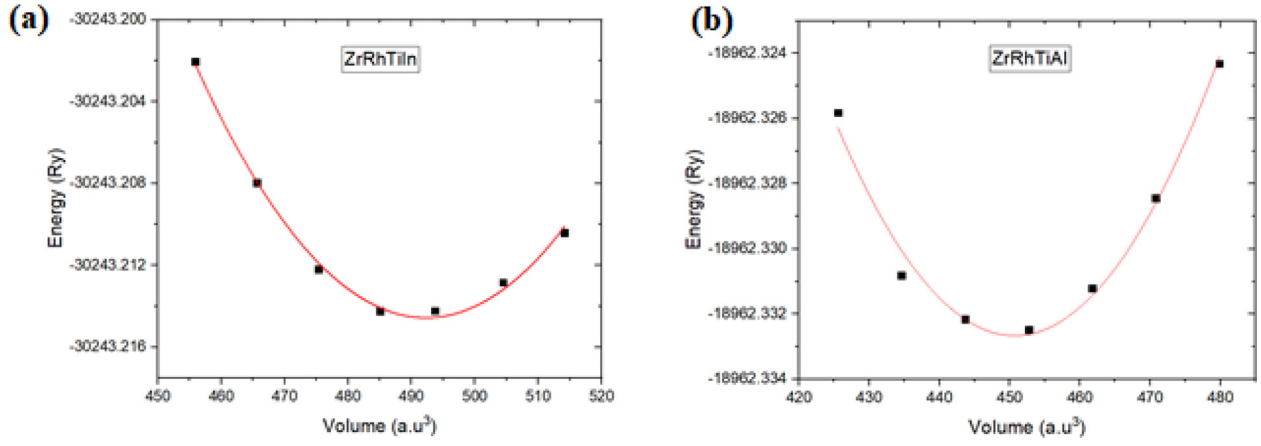


Fig. 1. (a) Energy vs Volume curve of ZrRhTiIn and (b) ZrRhTiAl.

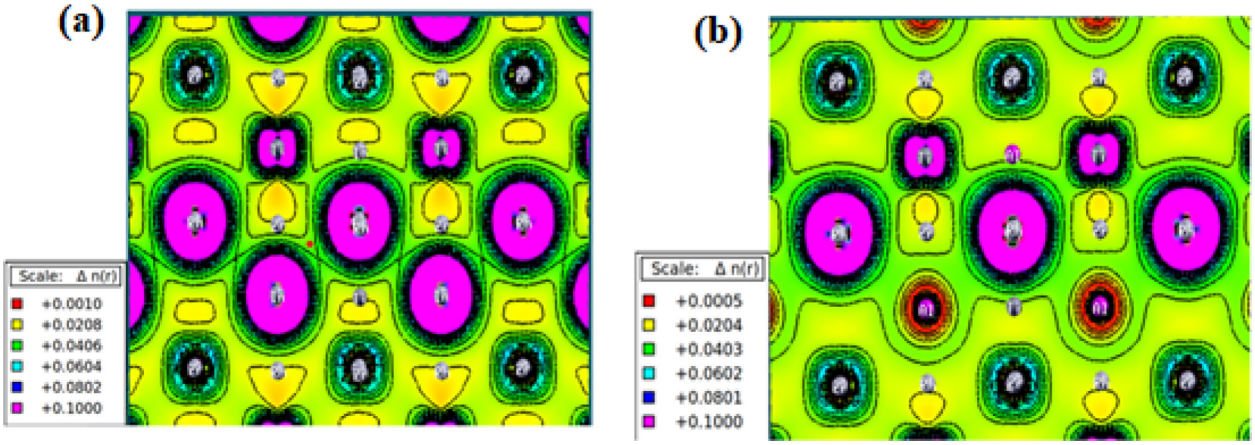


Fig. 2. Electronic charge density contour in (1 1 0) plane for (a) ZrRhTiIn and (b) ZrRhTiAl.

Table 1
Calculated elastic parameters for the ZrRhTiIn and ZrRhTiAl Heusler alloys.

Elastic parameters	ZrRhTiIn	ZrRhTiAl
C_{11} (GPa)	144.381	134.209
C_{12} (GPa)	118.126	128.502
C_{44} (GPa)	44.592	54.788
C_{11} - C_{12} (GPa)	27.552	12.737
C_{11} + $2C_{12}$ (GPa)	380.663	391.213
G (shear modulus) (GPa)	44.592	54.788
Cp (Cauchy's pressure) (GPa)	73.534	73.714
B (Bulk modulus) (GPa)	126.445	126.395
β (compressibility)	0.007908	0.007911
Pugh's ratio (B/G)	2.8355	2.306
A (Zener's value)	3.2369	8.602
GR (Reuss shear modulus) (GPa)	23.5342	13.557
GV (Voigt shear modulus) (GPa)	160.031	340.0142
GH (Hill shear modulus) (GPa)	91.7826	23.7858
E (young's modulus) (GPa)	221.704	67.1454
V (Poisson coefficient)	0.207	0.41146

$$Pugh's ratio = \frac{B}{G} \tag{11}$$

The (B/G) Pugh's ratio helps us to understand nature of the material whether it is ductile (greater than 1.75) or brittle (<1.75)[18]. The computed Pugh's ratio for ZrRhTiIn is 2.83, indicating that the material is ductile, whereas the Pugh's ratio for ZrRhTiAl is 2.30, indicating that the material is also ductile. The Cauchy pressure (C_{12}–C_{44}), which is positive for ductile materials and negative for brittle, supports this. Poisson ratio predicts the nature of the forces present, ν < 0.25 suggest the presence of non-central force, while 0.25 ≤ ν ≤ 0.50 indicates central force. The calculated Poisson ratio for ZrRhTiIn is 0.20 which suggests the presence of non-central force while it is 0.41 for ZrRhTiAl which suggests the presence of central force. Poisson ratio can also help in understanding more about the nature of bonding. If ν is near to or more than 0.25 then it signifies the ionic bonding while it is<0.25 it signifies the covalent bonding. Zener's value (A) is found to be 3.23 for ZrRhTiIn and 8.60 for ZrRhTiAl which shows that both the materials are anisotropic (other than unity) [19].

3.4. Electronic properties

We have plotted the spin-polarized band structure using the TB-mBJ potential. Both materials, ZrRhTiZ (Z = In, Al), behave as metal for the spin up channel and exhibit semiconducting nature in the spin down channel, with a band gap at the Fermi level. The value of this band gap is 0.39 eV for ZrRhTiIn and 0.30 eV for ZrRhTiAl. The valence band maximum and conduction band minimum are respectively situated at Γ and L, revealing a ZrRhTiZ (Z = In, Al) alloys have indirect band gap. (Figs. 3 and 4). The occurrence of a band gap at the fermi level in the down spin state implies that these alloys have 100 % spin polarization.

There are three possible explanations for the origin of the band gap in Heusler alloy. 1) - covalent bonding, 2) d-d hybridization, 3) - charge transfer. The hybridization of the Zr-4d and Rh-4d atoms forms 5 bonding bands (BB) and 5 non-bonding (NB) bands in the compounds, with 3 t_{2g} and 2 e_g in BB and 2 e_u and 3 t_u in NB. The 3d orbital of the Ti hybridised with 4d orbitals of 5 Zr-Rh BB forming the bonding and antibonding bands, while 5 NB 4d orbitals remain unhybridized. As a result, 15d orbitals are assigned to the spin-up and spin-down states, which are designated as 3 t_{2g}, 2 e_g, 2 e_u, 3 t_{1u}, 3 t_{2g}, 2 e_g, organised between higher and lower energy levels. The 1 s and 3p orbitals of Z (In, Al) are completely filled and are below the 15d hybridised orbitals. Because the triple degenerated t_{1u} states are vacant, a band gap forms between the t_{1u} (NB) and t_{2g} (BB) orbitals in the spin down state.

3.5. Transport property

The BoltzTraP code, which is based on the Semi-classical Boltzmann transport theory, was used to investigate the transport behaviour of these compounds by calculating various properties i.e., Seebeck coefficient, electrical and thermal conductivities, and power factor with temperature up to 1200 K under constant relaxation time (τ)[20]. The electrical conductivity and Seebeck coefficient are calculated using the generic Boltzmann transport equation [18,21], as shown in the formulae below;

$$\sigma_{\alpha,\beta}(\varepsilon) = \Sigma\tau v_\alpha(i,k)v_\beta(i,k)\frac{\delta(\varepsilon-\varepsilon_{i,k})}{\delta\varepsilon} \tag{12}$$

$$S = E_i\Delta jT^{-1} = (\sigma)\alpha^{-1}v_{\alpha,j} \tag{13}$$

Where τ, N and ε are relaxation time, number of k-points and band energy, respectively. α and β are electrical conductivity tensor indices and v_α (i,k) is group velocity.

3.5.1. Seebeck coefficient

The material with high Seebeck coefficient is considered to be a good for thermoelectric applications[22]. Fig. 5 (a–b) shows S as a function of temperature for current alloys. The obtained Seebeck Coefficient for the ZrRhTiAl is positive for the majority of temperature ranges, as shown in Fig. 5 (a), which helps us to deduce the fact that the alloy is a P-type semiconductor. The value of Seebeck

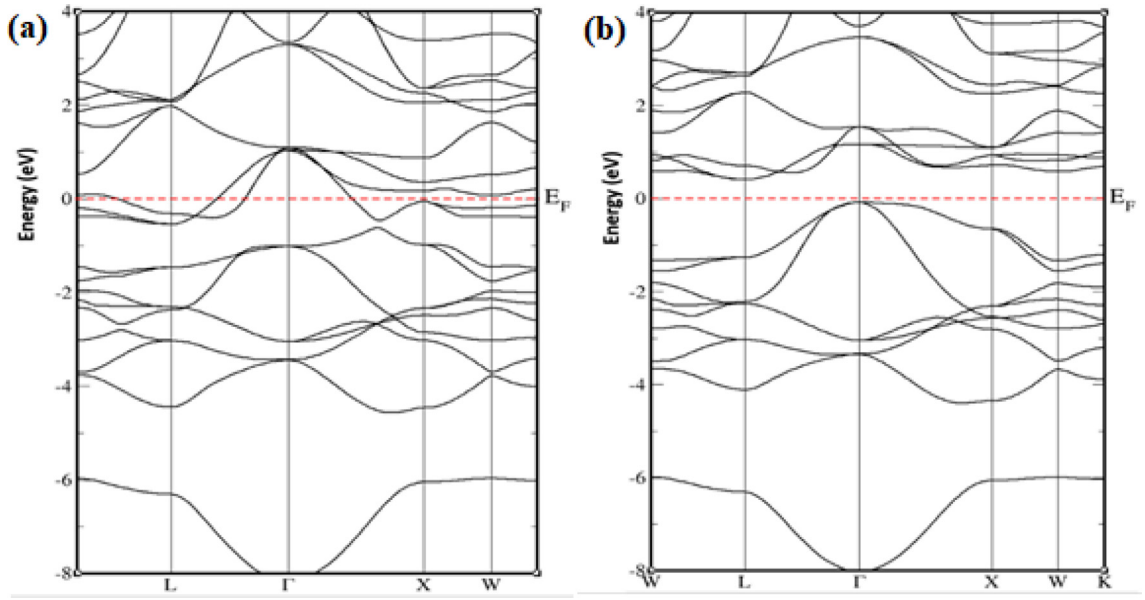


Fig. 3. Spin polarised band structure of ZrRhTiIn within the TB-mBJ approach for (a) spin up state and (b) spin down.

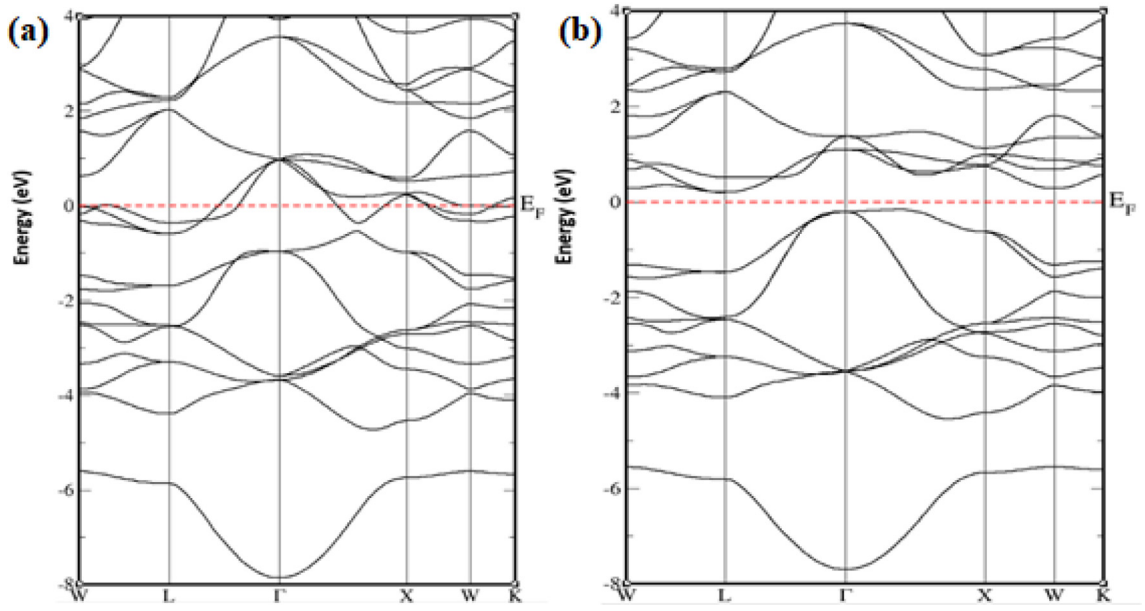


Fig. 4. Spin polarised band structure of ZrRhTiIAl within the TB-mBJ approach for (a) spin up state and (b) spin down.

Coefficient starts from $\sim 15.7\ \mu VK^{-1}$ at 150 K and starts falling till 550 K with value $\sim -1.83\ \mu VK^{-1}$ and again increases with temperature up to $\sim 4.01\ \mu VK^{-1}$ at 1200 K. At 300 K its value is $\sim 2.67\ \mu VK^{-1}$ which is not so good value as compared with other materials. Fig. 5 (b) shows the variation of Seebeck coefficient for ZrRhTiIn alloys and it can be seen that for most ranges of temperature, it is negative which suggests that this alloy is a N-type semiconductor. Its value starts from $\sim -101.2\ \mu VK^{-1}$ at 150 K, then decreases till $\sim -9.20\ \mu VK^{-1}$ at 350 K and then again goes into positive value with an increase in temperature up to 800 K with the value of 5.47 μVK^{-1}. Its value at room temperature is around $-25.40\ \mu VK^{-1}$ suggesting that it is better than the former for energy conservation.

3.5.2. Electrical conductivity (σ/τ)

The amount of electric charge or heat that may pass through the Fermi level is measured by electrical conductivity. Fig. 6(a-–b) depicts the fluctuation in electrical conductivity (σ) with temperature(T) for both alloys. From the Fig. 6 (a) we can see that electrical conductivity value first increases as temperature increases, from $\sim 65.6 \times 10^{20}\ (\Omega ms)^{-1}$ at 150 K to $\sim 73.0 \times 10^{20}\ (\Omega ms)^{-1}$ at 600 K and starts decreases till $\sim 71.7 \times 10^{20}\ (\Omega ms)^{-1}$ at 1200 K, the value of the electrical conductivity at room temperature is $\sim 71.23 \times 10^{20}\ (\Omega ms)^{-1}$ for ZrRhTiAl alloy. The electrical conductivity of the ZrRhTiIn alloy shows a roughly linear fluctuation with temperature, as shown in Fig. 6 (b). The value of electrical conductivity starts from $\sim 0.049 \times 10^{20}\ (\Omega ms)^{-1}$ at 150 K increases

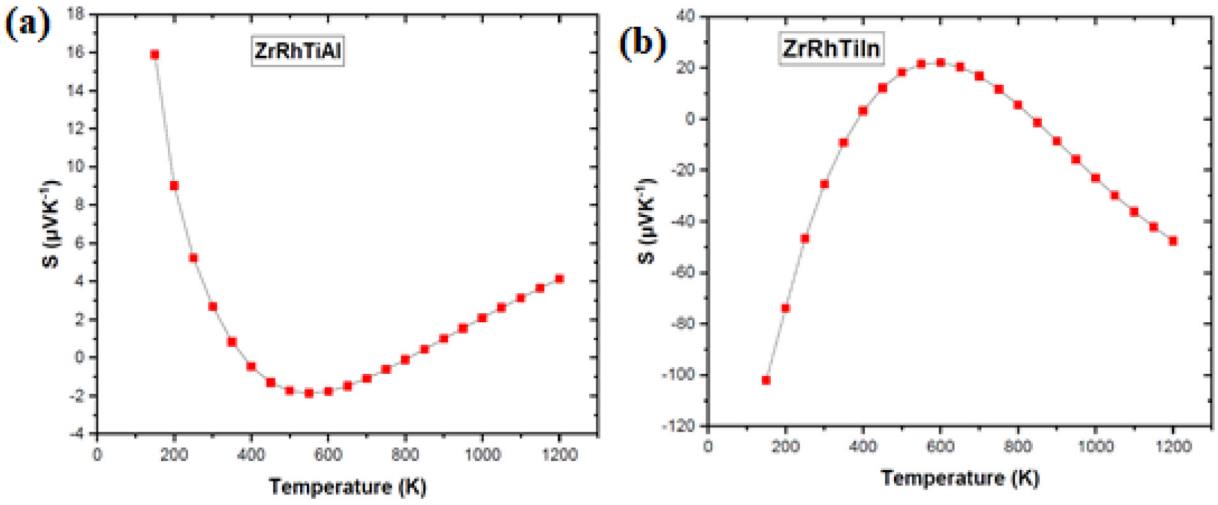


Fig. 5. Seebeck coefficient variation with temperature for (a) ZrRhTiAl and (b) ZrRhTiIn.

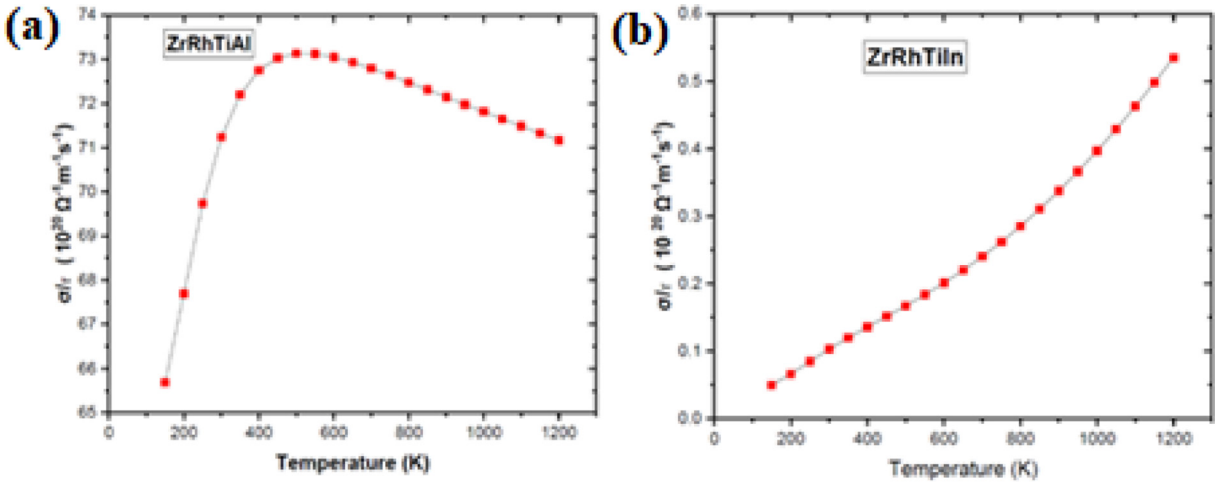


Fig. 6. electrical conductivity variation with temperature for (a) ZrRhTiAl and (b) ZrRhTiIn.

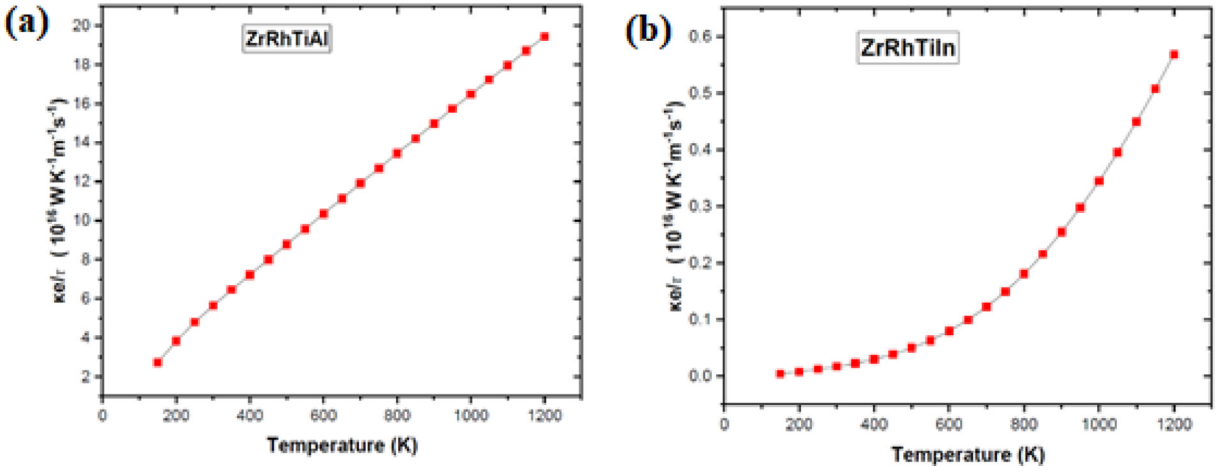


Fig. 7. Variation of electronic thermal conductivity with temperature for (a) ZrRhTiAl and (b) ZrRhTiIn.

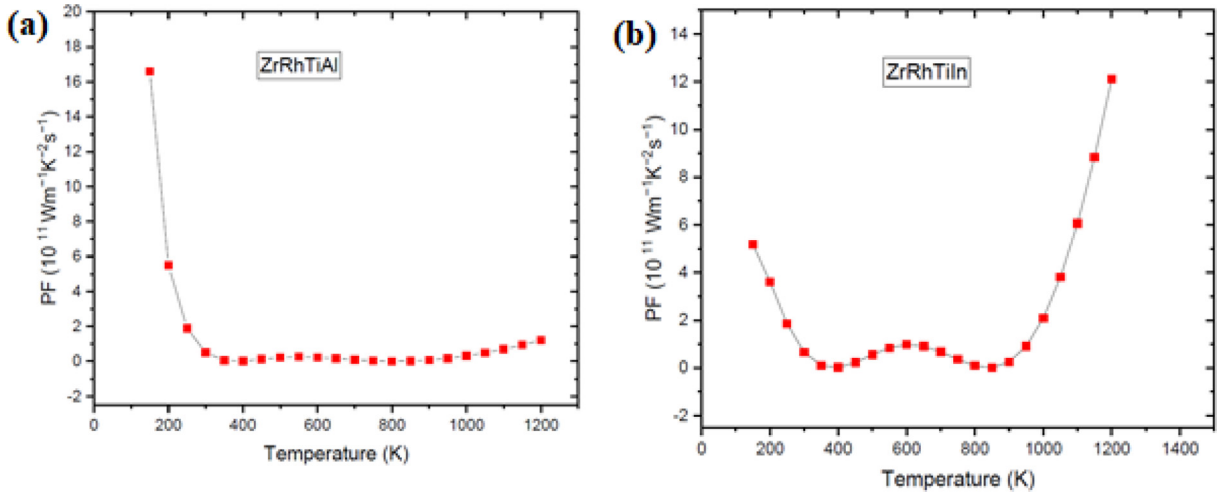


Fig. 8. Power factor variation with temperature for (a) ZrRhTiAl and (b) ZrRhTiIn.

as the temperature rises. The value of σ/τ at 1200 K is 0.535×10^{20}-$(\Omega ms)^{-1}$and the value of electrical conductivity for room temperature is $\sim 0.103 \times 10^{20}$ $(\Omega ms)^{-1}$.

3.5.3. Electronic thermal conductivity (κ_e/τ)

A material's thermoelectric performance is inversely proportional to its thermal conductivity. Therefore, for better thermoelectric efficiency, it should be very low[23]. Fig. 7(a–b) shows the variation in the electrical part of thermal conductivity with varying of temperature. Fig. 7 (a) depicting that, that the electronic thermal conductivity increases linearly from $\sim 2.72 \times 10^{16}$ (W $m^{-1}K^{-1}s^{-1}$) to 19.4×10^{16} ($Wm^{-1}K^{-1}s^{-1}$) for the ZrRhTiAl. The value of electronic thermal conductivity at room temperature is $\sim 5.64 \times 10^{16}$-($Wm^{-1}K^{-1}s^{-1}$). Similarly, Fig. 7(b) shows that it increases exponentially as temperature rises from 0.043×10^{16}($Wm^{-1}K^{-1}s^{-1}$) to 0.568×10^{16} ($Wm^{-1}K^{-1}s^{-1}$) with room temperature value $\sim 0.017 \times 10^{16}$ (W $m^{-1}K^{-1}s^{-1}$) for ZrRhTiIn. The ZrRhTiIn has lower value than ZrRhTiAl means that the former will be more useful as heat dissipation will be less in the ZrRhTiIn alloy.

3.5.4. Power factor

The power factor (PF) is a parameter which defines the material's efficiency for their thermoelectric application. The PF can be written as,

$$PF = S^2\sigma/\tau \quad (15)$$

σ/τ is the electrical conductivity, and S is the Seebeck coefficient[24]. Fig. 8 (a–b) plots the PF against the temperature for both the EQH's. Fig. 8 (a) shows the plot of PF against temperature from 150 K to 1200 K for ZrRhTiAl, it can be observed that the PF has maximum value of $\sim 17.74 \times 10^{11}$ $Wm^{-1}K^{-2}s^{-1}$ at 150 K, as temperature increases the PF decreases till 500 K having value $\sim 0.2101 \times 10^{11}$ $Wm^{-1}K^{-2}s^{-1}$and again increase slowly or become nearly constant. The value of PF at room temperature is observed to be $\sim 1.57 \times 10^{11}$ $Wm^{-1}K^{-2}s^{-1}$ for ZrRhTiAl alloy. Similarly, the fluctuation of PF as a function of temperature for the ZrRhTiIn alloy is shown in Fig. 8 (b). It can be seen clearly that the PF has a value of 5.56×10^{11} $Wm^{-1}K^{-2}s^{-1}$ at 150 K and the PF value at 1200 K is $\sim 12.11 \times 10^{11}$ $Wm^{-1}K^{-2}s^{-1}$. The value of PF at room temperature is observed to be $\sim 0.66 \times 10^{11}$ $Wm^{-1}K^{-2}s^{-1}$.

4. Conclusions

In this present paper, we examined the equiatomic quaternary Heusler alloys ZrRhTiIn and ZrRhTiAl using the density functional theory approach and semi-classical methods for its electronic, mechanical and thermoelectric properties. Both the alloys have found to be semiconducting in spin-down state with an indirect band gap. Because of the small band gap, they are useful for thermoelectric applications. The observed results for band gap agree with the previously reported results[25]. The Seebeck coefficient for ZrRhTiIn is -25.40 μVK^{-1} and for ZrRhTiAl is ~ 2.67 μVK^{-1} at 300 K, the corresponding Power factors are $\sim 0.66 \times 10^{11}Wm^{-1}$-$K^{-2}s^{-1}$ and $\sim 1.57 \times 10^{11}$ W $m-1K^{-2}s^{-1}$, respectively. Comparatively, ZrRhTiIn is a more promising thermoelectric material than the ZrRhTiAl. The obtained values are not very well adequate for application but can be improved with appropriate electron and hole doping.

CRediT authorship contribution statement

Vivek Kumar: Conceptualization, Investigation, Writing – original draft. **Mohit Kumar:** Investigation, Data curation, Validation. **Mukhtiyar Singh:** Supervision, Methodology, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgment

The authors wish thanks the University Grants Commission (UGC) of India for funding the computational facilities at the Department of Applied Physics, Delhi Technological University, Delhi, India, under the START-UP GRANT (No. F.30-416/2018 (BSR), dated 18/09/2018.

References

[1] X. Wang, Z. Cheng, J. Wang, X.-L. Wang, G. Liu, Recent advances in the Heusler based spin-gapless semiconductors Liu, J. Mater. Chem. C 4 (30) (2016) 7176–7192.

[2] T.M. Bhat, D.C. Gupta, Robust thermoelectric performance and high spin polarisation in CoMnTiAl and FeMnTiAl compounds, RSC Adv. 6 (83) (2016) 80302–80309.

[3] Lue C. S, C. F. Chen, J. Y. Lin, Y. T. Yu, and Y. K. Kuo, Thermoelectric properties of quaternary Heusler alloys Fe2 VAl1-x Six. *Phys. Rev. B* 2007;75;064204.

[4] L. Bainsla, K.G. Suresh, Equiatomic quaternary Heusler alloys: A material perspective for spintronic Applications, Appl. Phys. Rev. 3 (3) (2016) 031101.

[5] A. Kundu, S. Ghosh, R. Banerjee, S. Ghosh, B. Sanyal, New quaternary half-metallic ferromagnets with large Curie temperatures, Sci. Rep. 7 (1) (2017).

[6] H. Alqurashi, B. Hamad, Magnetic structure, mechanical stability and thermoelectric properties of VTiRhZ (Z = Si, Ge, Sn) quaternary Heusler alloys: first-principles calculations, Appl. Phys. A. 127 (10) (2021).

[7] M. Singh, H.S. Saini, S. Kumar, M.K. Kashyap, Effect of substituting sp-element on half metallic ferromagnetism in NiCrSi Heusler alloy, Comput. Mater. Sci. 53 (1) (2012) 431–435.

[8] M. Singh, H.S. Saini, J. Thakur, A.H. Reshak, M.K. Kashyap, Electronic structure, magnetism and robust half-metallicity of new quaternary Heusler alloy FeCrMnSb, J. Alloys. Compd. 53 (2013) 201–204.

[9] I. Galanakis, P.H. Dederichs, N. Papanikolaou, Slater-Pauling behavior and origin of the half-metallicity of the full-Heusler alloys, Phys. Rev. B. 66 (17) (2002).

[10] Q. Gao, H.H. Xie, L. Li, G. Lei, J.B. Deng, X.R. Hu, First-principle study on some new spin-gapless semiconductors: The Zr-based quaternary Heusler alloys, Superlattices Microstruct. 85 (2015) 536–542.

[11] S. Yousuf, D.C. Gupta, Investigation of electronic, magnetic and thermoelectric properties of Zr2Ni (Z = Al, Ga) ferromagnets, Mater. Chem. Phys. 192 (2017) 33–40.

[12] A.A. Mubarak, S. Saad, F. Hamioud, M. Al-Elaimi, Structural, thermo-elastic, electro-magnetic and thermoelectric attributes of quaternary CoNbMnX (X = Al, Si) Heusler alloys, Solid State Sci. 111 (2021) 106397.

[13] S.A. Sofi, D.C. Gupta, Systematic study of ferromagnetic phase stability of Co-based Heusler materials with high figure of merit: Hunt for spintronics and thermoelectric applicability, AIP Adv. 10 (10) (2020) 105330.

[14] S. Sharma, S.K. Pandey, Investigation of thermoelectric properties of half-metallic Co 2MnGe by using, first principles calculations, J. Phys. Condens. Matter. 26 (21) (2014) 215501.

[15] S. Ghosh, S. Ghosh, Half-Metallicity in Quaternary Heusler Alloys with 3d and 4d Elements: Observations and Insights from DFT Calculations, Phys. Status Solidi B. 256 (8) (2019) 1900039.

[16] S. Berri, Computational Study of Structural, Electronic, Elastic, Half-Metallic and Thermoelectric Properties of CoCrScZ (Z=Al, Si, Ge, and Ga) Quaternary Heusler Alloys, J. supercond. Nov. Magn. 33 (12) (2020) 3809–3818.

[17] H. Moujri, M. Berber, M. Mebrek, A. Boudali, T. Ouahrani, Structural, elastic, electronic, and magnetic properties of new quaternary Heusler alloy PdCoMnGa and PdCoMnAl. *Indian*, J. Phys. 96 (4) (2022) 1025–1035.

[18] S. Ahmad Khandy, J.-D. Chai, Robust stability, half-metallic ferrimagnetism and thermoelectric properties of new quaternary Heusler material: A first principles approach, J. Magn. Magn. 502 (2020) 166562.

[19] Khandy S. A and J. da Chai, Thermoelectric properties, phonon, and mechanical stability of new half-metallic quaternary Heusler alloys: FeRhCrZ (Z = Si and Ge). *J. Appl. Phys.* 2020;127;165102.

[20] A.Q. Seh, D.C. Gupta, Comprehensive DFT investigation of transition-metal-based new quaternary Heusler alloys CoNbMnZ (Z = Ge, Sn): Compatible for spin-dependent and thermoelectric applications, RSC Adv. 10 (71) (2020) 43870–43881.

[21] Khandy S, A. I. Islam, D.C Gupta, M. A. Bhat, S. Ahmed, T. A Dar, S. Rubab, S. Dhiman and A. Laref, A case study of Fe2TaZ (Z = Al, Ga, In) Heusler alloys: hunt for half-metallic behaviour and thermoelectricity. *RSC Adv*. 2018;8;40996-41002.

[22] A.Q. Seh, D.C. Gupta, Exploration of highly correlated Co-based quaternary Heusler alloys for spintronics and thermoelectric applications, Int. J. Energy Res. 43 (2019) 8864–8877.

[23] S. Singh, D.C. Gupta, Lanthanum based quaternary Heusler alloys LaCoCrX (X = Al, Ga): Hunt for half- metallicity and high thermoelectric efficiency, Results in Phys. 13 (2019) 102300.

[24] T.T. Lin, Q. Gao, G.D. Liu, X.F. Dai, X.M. Zhang, H.B. Zhang, Dynamical stability, electronic and thermoelectric properties of quaternary ZnFeTiSi Heusler compound, Curr. Appl. Phys. 19 (6) (2019) 721–727.

[25] R.B. Ray, G.C. Kaphle, R.K. Rai, D.K. Yadav, R. Paudel, D. Paudyal, Strain induced electronic structure, and magnetic and structural properties in quaternary Heusler alloys ZrRhTiZ (Z = Al In), *J. Alloys compd.* 867 (2021) 158906.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Investigations on effect of bending radius on springback behaviour of three-ply clad sheet

Pankaj Kumar Sharma [a,b,*], Vijay Gautam [a], Atul Kumar Agrawal [a]

[a] Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[b] Department of Mechanical Engineering, Raj Kumar Goel Institute of Technology, Ghaziabad, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Springback
3-ply clad sheet
Bending radius
Analytical
Numerical
and Experimental results

ABSTRACT

Bending operation is widely used in manufacturing of automobiles, aircraft, spacecraft, structures, and home appliances etc. During bending operations, springback is a very common and critical defect. For dimensional accuracy springback must be minimized sufficiently. In the present investigation, the effect of bending radius/punch profile radius on the springback behaviour of SS304/Al1050/SS304 clad sheet has been studied. It has been observed that bending radii /punch profile radii have great impact on the prediction of springback of clad sheet. Springback can be reduced by reducing the bending radius/punch profile radius. The present study confirms that the springback can fairly be controlled by taking optimal value of bending radius/punch profile radius. Sheet placing arrangement of 3-ply clad sheet also affect the springback behaviour minutely. A good agreement has been observed between analytical, numerical, and experimental results of the springback of a 3-ply clad sheet. To evaluate the mechanical properties of three-ply clad sheet, tensile and micro hardness testing have been conducted. The average hardness of 3-ply clad sheet is evaluated as 113.18 HV which may be attributed to the individual hardness of layers of SS304, Al1050 and SS304.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Recent Advances in Modeling and Simulations Techniques in Engineering and Science.

1. Introduction

Metal laminated sheet composites are a unique form of composite material in which alternating metal or metal containing layers are bonded together with inter-metal bonding. Various types of metal laminated sheet are widely applying in several applications such as spacecraft, aircraft, automotive, fabrication, electrical and structures units as per their specific and unique properties. To design and develop the laminated composite materials comprising of the optimum level of various properties involves the stacking layers of various thicknesses of different materials, interface structures and properties of different materials to be affixed together [1,2]. There are many solid-state bonding techniques like cold roll bonding (CRB), diffusion and explosive bonding for manufacturing of laminated sheets/clad sheets but CRB (cold roll bonding) is very cost effective and most efficient technique [1,3]. Basically, cold roll bonding process is a solid phase welding process, in which intimate bond between the sheets is established by plastic deformation of the sheet metals [4]. Engineers and researchers are showing their keen interest in the lightweight composite materials and reason on behind this is to their unique, usable and optimal physical properties which are responsible for the remarkable developments in the forming behaviour of the metal laminated sheets. However, press forming of the components of clad sheet metals is a difficult task because their elasto-plastic behaviour is completely different from that of their individual part sheets. The plastic deformation of a laminated sheet lies on the properties of individual materials, sequence of placing of the different sheets, thickness ratio, surface related properties, bonding techniques & their parameters and subsequent heat treatment processes [1,3,5,6]. Especially, their bending and springback characteristics are completely different from those of monolithic sheets resultant from the inherent modification in their mechanical properties and change in thickness direction [7].

* Corresponding author at: Department of Mechanical Engineering, Delhi Technological University, Delhi, India.
E-mail addresses: erpankajsharma719@gmail.com, panmefme@rkgit.edu.in, erpankajsharma719@gmail.com, panmefme@rkgit.edu.in (P. Kumar Sharma).

https://doi.org/10.1016/j.matpr.2022.04.601
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Recent Advances in Modeling and Simulations Techniques in Engineering and Science.

Please cite this article as: P. Kumar Sharma, V. Gautam and A. Kumar Agrawal, Investigations on effect of bending radius on springback behaviour of three-ply clad sheet, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.601

The most popular metal laminated sheets are Ti/ferritic stainless steel, Al/ferritic stainless steel, Al/Cu, Al/Fe, AZ31/Al/ferritic stainless steel, Al/Mg/Al, Ti/steel/Ti and austenitic stainless steel/Al/Cu [5–8].

Laminated composites become effective and functional when different types of metals with distinct metallic properties are combined into a single sheet. The composition of the laminated composites is framed and designed in such a way that shortcomings of other materials can be makeup, simultaneously utilizing the advantageous characteristics of each metal which, as a whole, provides a unique sophisticated function [9].

3-ply, SS304/Al1050/SS304 clad composites are getting popularity among the laminated composites as they are widely using in manufacturing of cook wares, automotive parts, electric components, rail transport and aerospace related components due to their important mechanical properties like higher strength to weight ratio, higher electrical & thermal conductivity, better ductility & formability, resistance to scratch and corrosion. In the manufacturing of different parts, several bending operations are performed. During bending operations, springback is a very common and critical defect. For dimensional accuracy and achieving high precisions springback must be in a controlled manner. Many researchers have studied springback of monolithic materials but there are very few who have worked in the field of springback of clad sheets. In this research paper, investigations on effect of bending radius on springback behaviour of 3-ply clad sheet (SS304/Al1050/SS304) in V-bending has been done which never done before. This is a novel work.

2. Materials and experimental procedures

2.1. Materials of 3-Ply laminated sheet

A 3-ply laminated sheet which is manufactured by cold roll bonding (CRB) process containing layers of SS304, AA1050 and SS430 having thickness 0.4 mm, 1.4 mm, and 0.6 mm respectively, has been used for tensile testing, hardness testing and determination of anisotropy & sprinback. Thus, the total thickness of 3-ply sheet is 2.4 mm. AA1050 layer is laminated as a sandwich in between the layers of austenitic SS304 and ferritic SS430 stainless steels. High strength of the SS304 and SS 430 layers enhance the overall strength of the 3-ply sheet.



Fig. 1. Tensile samples at different orientations after testing.

2.2. Tensile testing

To obtain the tensile properties of 3-ply sheet metal, the samples for uniaxial tension test are cut by CO2-laser in the orientation of 0-degree, 45°, and 90°with respect to the rolling direction of the laminated sheet as shown in Fig. 1. All the samples are cut and prepared according to ASTM-E8M. The tensile samples of the steel layers are obtained from 3-ply by dissolving aluminum using a solution of NaOH. Complete dissolution of AA1050 takes one month but this time can be decreased to two to three weeks by increasing the temperature and concentration of the NaOH solution. To obtain the samples of AA1050 alone from the 3-ply sheet, the layer of stainless steel is removed by fine-controlled facing operation on a full-sized blank to a predetermined depth equal to the thickness of the SS430 and SS304.

All the tensile and anisotropy tests are performed on 50 kN capacity of UTM with a crosshead speed of 2.5 mm/min. Each tensile test is performed thrice to capture the reproducibility of the experimental data points. The engineering stress–strain curves are transformed into plots of true stress–true strain curves. The log–log plots of true stress–true strain gives the values of strain hardening exponent (n) and strength coefficients (K).

2.3. Micro hardness testing

Struers hardness tester (DURAMIN-40 M1) has been used to obtain the micro hardness of SS304/AA1050/SS430 clad sheet. In this testing, with the help of a diamond indenter, a mark (indentation) is made on the sample by applying a suitable load. The size of indentation is measured with the help of a calibrated optical microscope, and the hardness is evaluated as the mean stress applied underneath the indenter. With the help of a diamond indenter three, seven and four indentations have been made on SS304, AA1050 and SS430 portion of laminated clad sheet and further hardness values have been evaluated with respect to these indentations as shown in Fig. 2.

2.4. Experimental procedure for determination of springback

To obtain values of sprinback of 3-ply sheets in V-bending, 50 kN UTM ((Tenius Olsen) has been used. To perform V- bending, different punch and die sets are employed. All punches and dies are manufactured by wire cut electric discharge machine using D2 steel. After completing manufacturing of the tool & die, they have undergone by a suitable heat treatment process. The punch and dies sizes are taken 15 mm, 17.5 mm, 20 mm and 17.4 mm, 19.9 mm, 22.4 mm respectively. To restrict any localized compression of 3-ply sheet, 2.4 mm clearance has been employed. In all the cases, it has been confirmed that punch will be always fixed in the movable jaw and die will be fixed into stationary jaw of 50 kN capacity of UTM as shown in Fig. 3.

To find out the springback experimentally, two setting conditions have taken according to sheets placements: -.

1. SS430 as inner layer (contacting with punch surface) and SS304 as outer layer (contacting with punch surface).
2. SS304 as inner layer (contacting with punch surface) and SS430 as outer layer layer (contacting with punch surface).

In the first case when SS430 is considered as an inner layer, the values of bending radii are 15.7495 mm, 18.2495 mm and 20.7495 mm but in the second case when SS304 is considered as an inner layer, the values of bending radii are 16.2505 mm, 18.7505 mm and 21.2505 mm against the different punch profile radii of 15 mm, 17.5 mm, and 20 mm respectively. The values of

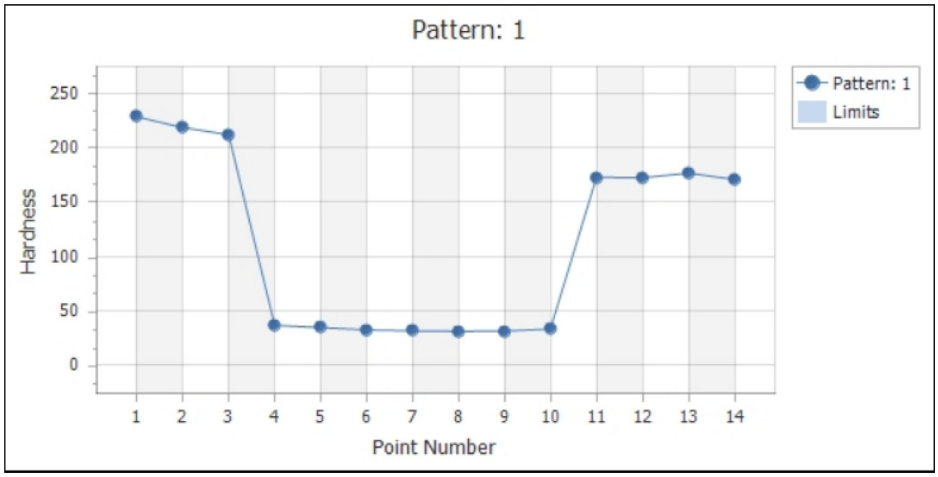


Fig. 2. Micro Hardness of the 3-ply clad sheet at various location of entire thickness.

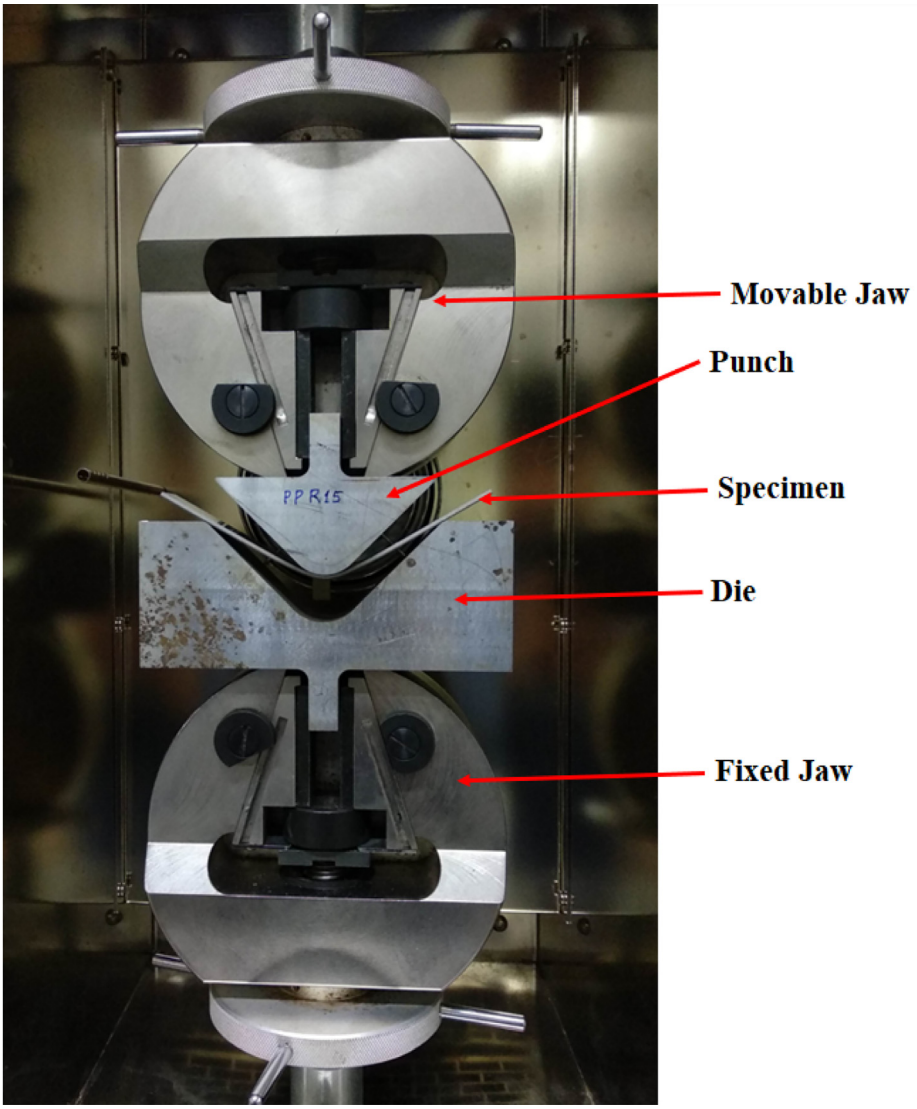


Fig. 3. V-bending experimental setup.

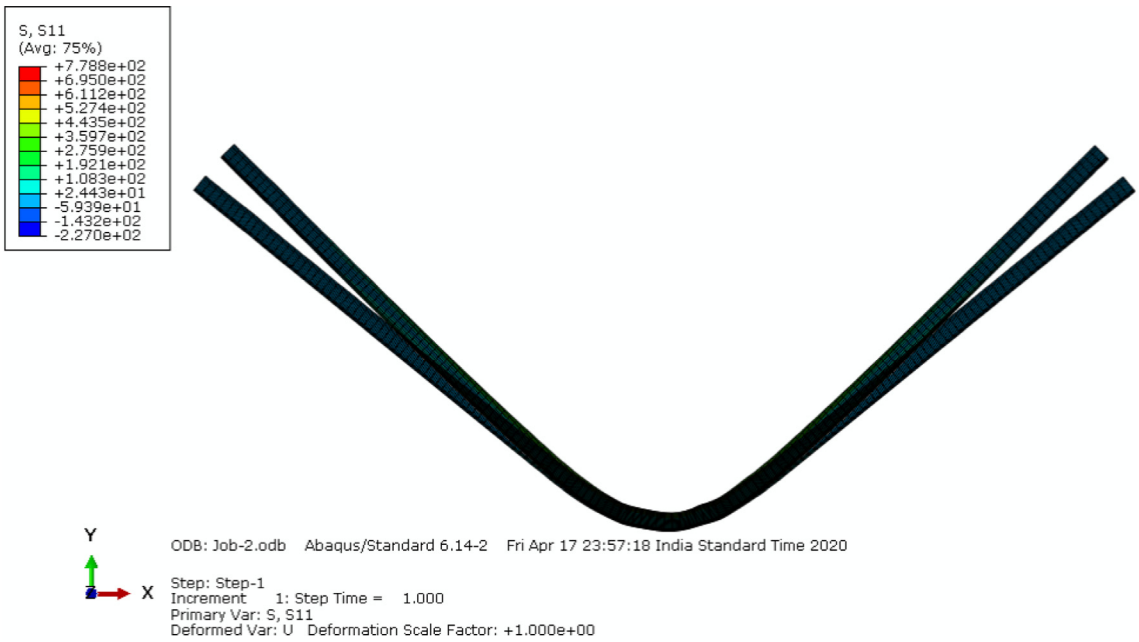


Fig. 4. Overlay plot of three-ply clad sheet before and after sprinback.

bending radii are more than in the second case as compared to the first case.

The bending samples size is selected as 150mmX25mm and cut by a wire cut electric discharge machine (EDM) from 3-Ply clad sheet. After completing the bending experiments, included bend angles of all the samples of 3-ply sheet have been measured by using sensor-based vision inspection machine.

2.5. Determination of springback by FE simulations

ABAQUS software of version 6.14–2 has been used to obtain the springback after bending of the 3-ply clad sheet of SS430/AA1050/SS304. Punch and die set modelled as rigid body so that time of simulation can be decreased when the bending sample is modelled as deformable with homogeneous solid shell elements. The modeled geometry of the blank is divided longitudinally into three regions of SS 304 layer of thickness 0.4 mm, AA1050 layer of thickness 1.4 mm and SS430 layer of thickness 0.6 mm using a partition tool in the software. The partitioned layer of SS304 and SS430 are meshed finer due to their lower thickness as compared with a layer of AA1050. The specimen for bending is modeled as deformable with homogeneous solid shell elements, i.e., CPE4R plane-strain elements which are four node bilinear, reduced integration with hourglass control. True stress–true strain data points after plastic yielding for individual materials obtained by uniaxial tension tests of AA1050, SS430 and SS304 are used to assign the specific plastic properties of all the materials to the blank. The bending samples are placed on the die in two ways, i.e., in the first case, SS430 is the inner layer and SS304 is the outer layer whereas in the second case, SS304 is the inner layer and SS430 is the outer layer. The simulations are done for bending and springback for both the cases of sheet placements on the die to assess the effect of the inner and outer layer on bending and springback. In the assembly module, the blank which will be further bent is put on the die in such a manner that the top surface of the blank touches punch and considered as inner side. The surfaces of the die and the punch are represented as master surfaces and the surface of the blank represented as slave surface. The friction between die and blank interface and punch and blank interface are taken as 0.125 and 0.05 respectively. The punch is set with intimate contact with the blank according to position of the inner layer of the blank. The displacement of the punch is given equally to the die depth.

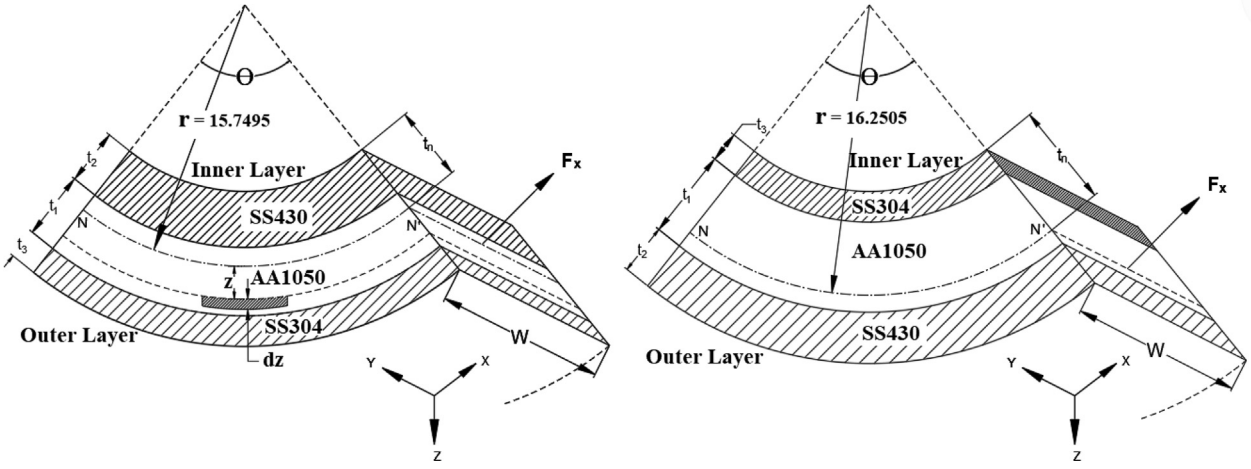


Fig. 5. Two different bending radii for same punch profile radius (PPR) of 15 mm.

Die is kept fixed during the whole displacement of punch towards die. All the bending simulations for both the sheet placements on the die are performed using dynamic explicit. To evaluate and capture the results of spring back simulations, the initial output file from the simulated bending process history is utilized after deleting all the constraints in the model of springback. Fig. 4 shows an overlay plot of loaded and unloaded images after the spring back simulation of a specimen. The coordinate details of the various nodes are captured individually from the loaded and unloaded frames for springback measurement. The final value of springback can be determined by computing the difference between the two carefully recorded coordinates of both the frames in a computer aided engineering (CAE) interface [10–12].

2.6. Prediction of springback by analytical method

The existing analytical model for prediction of springback in bending of 3-ply clad sheet incorporating the effects of strain hardening and anisotropy [11], is used for determining the springback in bending of 3-ply clad sheet SS430/Al1050/SS304 for different punch profile radii/bending radii. Punch profile radii have been taken 15 mm, 17.5 mm and 20 mm.

To find out the springback of 3-ply clad sheet, two setting conditions have taken according to sheets placements as shown in Fig. 5: -.

1. SS430 as inner layer (contacting with punch surface) and SS304 as outer layer (contacting with die surface).
2. SS304 as inner layer (contacting with punch surface) and SS430 as outer layer layer (contacting with die surface).

The distance of the neutral axis $'t_n'$ from the bottommost plane of the sheet considering the equivalent area is given as below:

$$t_n = \left[\frac{(2t_2+t_1)E_1t_1 + t_2^2E_2 + (2t_1+2t_2+t_3)E_2t_3}{2(E_1t_1+E_2t_2+E_3t_3)}\right]$$

Case-I: Bending of three-ply sheet with SS430 as the inner layer.

To obtain springback in both the cases of sheet placements i.e., Case-I and II, the final radius of curvature (r') after the springback is found by Eq. (1) and (2), respectively.

$$\text{Change in the bend angle, } \alpha_i - \alpha_f = \left(1 - \frac{r}{r'}\right)\alpha_i \tag{3}$$

where, α_i and α_f are initial and final bend angles. The included bend angle is 90° and is measured on the die [11].

3. Results and discussion

3.1. Tensile properties of three-ply sheet

The tensile properties of a 3-ply laminated sheet (SS430/AA1050/SS304) along 0-degree, 45-degree and 90-degree w.r.t rolling direction are represented in Table 1. The strain hardening exponent n, strength coefficient K, yield and tensile strength of the sample oriented at 90-degree to the rolling direction are higher than the samples having orientation 0-degree and 45-degree w.r.t. the rolling direction. The strength of the 3-ply laminated sheet depends on the individual strength of SS430, AA1050 and SS304 layers. The highest % elongation is seen in the sample which is oriented at 0-degree w.r.t. rolling direction. The normal anisotropy of the 3-ply laminated sheet (SS430/AA1050/SS304) is obtained to be 0.81 which is less than one.

3.2. Micro hardness

The average micro hardness of SS430 and AA100 are 173.5 HV and 32.91 HV respectively as per Table 2. The average micro hardness of SS304 is 220.05 HV which is more than the hardness of SS430 and AA1050. The average hardness of 3-ply clad sheet is evaluated as 113.18 HV. AA1050 layer is softest among all the layers of 3-ply clad sheet. The size of the impression of hardness tester is biggest and lowest on AA1050 layer and SS304 respectively. It indicates that AA1050 is the softest and SS304 is hardest material in the 3-ply clad sheet. The hardness of SS430 layer lies in between AA1050 and SS304 layers.

As per Fig. 6, the impression of hardness tester in SS430 layer and SS304 layer is fine, sharp, and small as compared to the layer of AA1050.

$$\left(\frac{1}{r}-\frac{1}{r'}\right) = \frac{\left[\frac{2w.E_1'.z_e^3}{3r} + \frac{w.K_1'.C_2^n}{C_1.r^n.(n+2)}\cdot\left[(t_1+t_2-t_n)^{(n+2)} - 2(z_e)^{(n+2)} + (t_n-t_2)^{(n+2)}\right] + \frac{K_2'.C_2^n.w}{r^nC_1.(n+2)}\left[(t_n)^{(n+2)} - (t_n-t_2)^{(n+2)}\right] + \frac{K_3'.C_2^n.w}{r^nC_1.(n+2)}\left[(t_1+t_2+t_3-t_n)^{(n+2)} - (t_1+t_2-t_n)^{(n+2)}\right]\right]}{E_1'.\left[\left(\frac{wt_1^3}{12} + \left((wt_1).(t_2+\frac{t_1}{2}-t_n)^2\right)\right).\right] + E_2'.\left[\left(\frac{wt_2^3}{12} + \left((wt_2).(t_n-\frac{t_2}{2})^2\right)\right).\right] + E_3'.\left[\left(\frac{wt_3^3}{12} + \left((wt_3).(t_1+t_2+\frac{t_3}{2}-t_n)^2\right)\right)\right]} \tag{1}[11]$$

Eq. (1) gives the final radius of curvature (r') after springback.

Case-II: Bending of three-ply sheet with SS304 as the inner layer.

$$\left(\frac{1}{r}-\frac{1}{r'}\right) = \frac{\left[\frac{2w.E_1'.z_e^3}{3r} + \frac{K_1'.C_2^n.w}{r^nC_1.(n+2)}\left[(t_1+t_3-t_n)^{(n+2)} - 2(z_e)^{(n+2)} + (t_n-t_3)^{(n+2)}\right] + \frac{K_2'.C_2^n.w}{r^nC_1.(n+2)}\left[(t_1+t_2+t_3-t_n)^{(n+2)} - (t_1+t_3-t_n)^{(n+2)}\right] + \frac{K_3'.C_2^n.w}{r^nC_1.(n+2)}\left[(t_n)^{(n+2)} - (t_n-t_3)^{(n+2)}\right]\right]}{E_1'.\left[\left(\frac{wt_1^3}{12} + \left((w.t_1).(t_n-\frac{t_1}{2}-t_3)^2\right)\right)\right] + E_2'.\left[\left(\frac{wt_2^3}{12} + \left((wt_2).(t_1+t_3+\frac{t_2}{2}-t_n)^2\right)\right) + E_3'.\left[\left(\frac{wt_3^3}{12} + \left((wt_3).(t_n-\frac{t_3}{2})^2\right)\right)\right]\right]} \tag{2}[11]$$

Table 1
Tensile properties of a clad sheet [11].

Sample's orientation w.r.t. rolling direction	Tensile Strength (MPa)	0.2% Yield Strength (MPa)	% Elongation	Strength coefficient (MPa) K	Strain hardening exponent n	Plastic strain ratio *R*
0°	219.5 ± 2.12	153.5 ± 0.71	49.50 ± 4.80	400.0 ± 2.83	0.255	0.64
45°	228.0 ± 1.41	174.0 ± 1.00	48.85 ± 1.48	398.5 ± 20.5	0.232	0.89
90°	237.5 ± 2.12	176.0 ± 1.41	44.45 ± 0.29	437.5 ± 4.95	0.266	0.83

Table 2
Micro Hardness of individual layers of the 3-ply clad sheet.

S.No.	Material	Hardness (HV)	Mean Hardness (HV)
1	SS304	228.90	220.05
2		218.91	
3		212.35	
4	AA1050	36.54	32.91
5		34.41	
6		32.41	
7		31.50	
8		31.01	
9		31.30	
10		33.19	
11	SS430	173.22	173.5
12		172.70	
13		176.95	
14		171.14	

3.3. Springback analysis

The springback results for various bending radii after the bending of 3-ply sheet by experiments, numerical simulations, and existing analytical model are given in Table 3 and Table 4. Highest springback values are observed in SS430 and SS304 which could be attributed to its highest strength and smaller sheet thickness. Lowest springback values are observed in AA1050 which could be attributed to its lower strength and higher sheet thickness. The specimens of 3-ply sheet oriented at 90 deg to the rolling direction tested with SS304 as inner layer experiences highest springback followed by the specimens oriented at 0 deg and 45 deg with respect to the rolling direction of the clad sheet which again may be attributed to the tensile strength of the sheet and the combined effects of anisotropy of all the layers at the zone of deformation. Highest springback values are obtained for 20 mm punch profile radius when lowest springback values are obtained for 15 mm punch profile radius. Similar results are observed in the springback values predicted by simulation and analytical models in all the cases. Springback values are increased due to increment in the

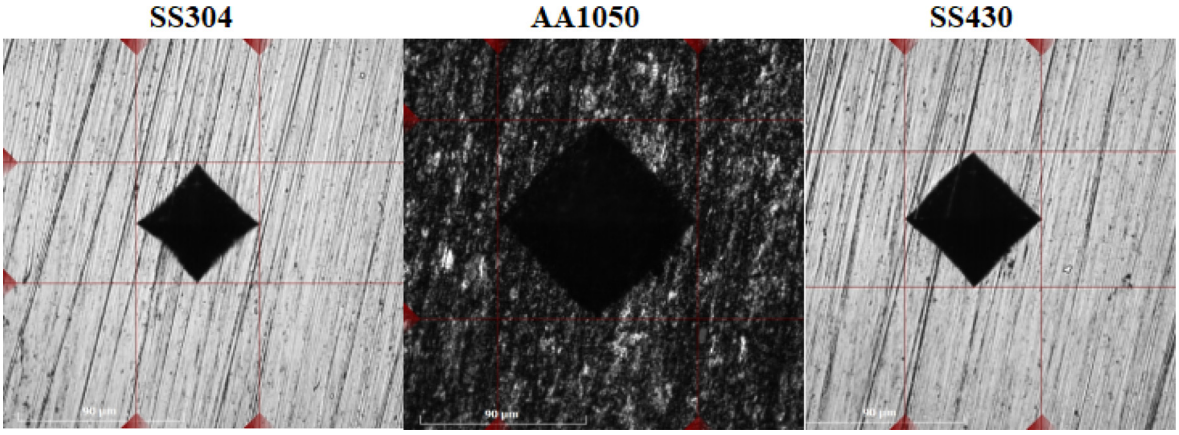


Fig. 6. Impressions of hardness tester (DURAMIN-40 M1) in SS430/AA1050/SS304 Clad Sheet.

Table 3
Numerical and Experimental springback results of a 3-ply sheet for varying bending radius.

Inner layer during bending		Bending radius 15.7495 mm/ PPR 15 mm			Bending radius 18.2495 mm/ PPR 17.50 mm			Bending radius 20.7495 mm/ PPR 20 mm		
		Exp.	Simulation	% Error	Exp.	Simulation	% Error	Exp.	Simulation	% Error
		In Degree			In Degree			In Degree		
SS430	0°	4.45	4.75	6.74	4.81	5.07	5.41	5.3	5.55	5.11
	45°	4.94	5.34	8.10	5.32	5.59	5.08	5.8	6.09	5.36
	90°	5.05	5.44	7.72	5.48	5.71	4.20	6	6.32	4.98
	Bending radius 16.2505 mm/ PPR 15 mm				Bending radius 18.7505 mm/ PPR 17.50 mm			Bending radius 21.2505 mm/ PPR 20 mm		
SS304	0°	4.51	4.67	3.55	4.89	5.17	5.73	5.4	5.61	4.08
	45°	5.01	5.21	3.99	5.41	5.74	6.10	6	6.19	3.86
	90°	5.15	5.39	4.66	5.61	5.84	4.10	6.1	6.45	5.56

Table 4
Analytical and Experimental springback results of a 3-ply sheet for varying bending radius.

Inner layer during bending		Bending radius 15.7495 mm/ PPR 15 mm			Bending radius 18.2495 mm/ PPR 17.5 mm			Bending radius 20.7495 mm/ PPR 20 mm		
		Exp.	ANA	% Error	Exp.	ANA	% Error	Exp.	ANA	% Error
		In Degree			In Degree			In Degree		
SS430	0°	4.45	4.06	8.76	4.81	4.50	6.44	5.3	4.91	7.01
	45°	4.94	4.44	10.12	5.32	4.93	7.33	5.8	5.40	6.57
	90°	5.05	4.60	8.91	5.48	5.08	7.30	6	5.55	7.81
		Bending radius 16.2505 mm/ PPR 15 mm			Bending radius 18.7505 mm/ PPR 17.5 mm			Bending radius 21.2505 mm/ PPR 20 mm		
SS304	0°	4.51	4.08	9.53	4.89	4.52	7.57	5.4	4.93	8.53
	45°	5.01	4.47	10.78	5.41	4.96	8.32	6	5.42	9.06
	90°	5.15	4.63	10.10	5.61	5.10	9.09	6.1	5.58	8.67

punch profile radius. Springback can be minimized by taking small punch radius. Sheet setting conditions affect springback a lot. Almost in all the cases higher springback values are obtained when SS304 is inner layer during bending which can attributed due to higher bending radius and high strength of SS304 as compared to SS430.

4. Conclusions

SS304 and SS430 layer are accountable for high strength and ductility of SS304/Al1050/SS304 clad sheet. In the current investigation it has been observed that the springback values are decreasing when bending radii/punch profile radii are decreased. The springback of SS304/Al1050/SS304 clad sheet can be minimized by taking lower bending radius/punch profile radius. Sheet placing pattern influenced the springback behaviour of 3-ply clad sheet. In maximum cases regarding springback experiments of samples of SS304/Al1050/SS304 clad sheets labelled as SS430_out/SS304_in, the springback values are higher than the values of springback obtained with the samples labelled as SS430_in/SS304_out which can be attributed due to more bending radius. No thickness change is observed in overall thickness of 3-ply clad sheet during bending due to strong protection of soft layer of AA1050 by high strength SS430 and SS304 layers. A good agreement has been established between analytical, numerical, and experimental results of the springback of a 3-ply clad sheet. But however numerical simulation results are near to the experimental results as compared to the analytical results which can be attributed due to assumptions made in the analytical model.

Credit author statement:

This research has been done by Pankaj Kumar Sharma under the guidance of Dr. Vijay Gautam, Professor in Department of Mechanical Engineering, Delhi Technological University, New Delhi and Dr. Atul Kumar Agarwal, Professor in Department of Mechanical Engineering, Delhi Technological University, New Delhi.

Pankaj Kumar Sharma has written this research paper titled on ‘‘Investigations on Effect of Bending Radius on Springback Behaviour of Three-Ply Clad Sheet” and Dr. Vijay Gautam and Dr. Atul Kumar Agarwal have edited this paper.

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.

References

[1] I.K. Kim, S.I. Hong, Effect of component layer thickness on the bending behaviors of roll-bonded tri-layered Mg/Al/STS clad composites, Mater. Des. 49 (2013) 935–944.
[2] V. Gautam, P. Sharma, K.D. Ravi, Experimental and Numerical Studies on Spring back in U-Bending of 3-Ply Cladded Sheet Metal, Mater. Today:. Proc. 5 (2018) 4421–4430.
[3] H.D. Manesh, A.K. Taheri, Bond strength and formability of an aluminum-clad steel sheet, J. Alloy. Compd. 361 (2003) 138–143.
[4] R. Jamaati, M.R. Toroghinejad, Effect of friction, annealing conditions and hardness on the bond strength of Al/Al strips produced by cold roll bonding process, Mater. Des. 31 (2010) 4508–4513.
[5] I.K. Kim, S.I. Hong, Mechanochemical joining in cold roll-cladding of tri-layered Cu/Al/Cu composite and the interface cracking behavior, Mater. Des. 57 (2014) 625–631.
[6] T. Mori, S. Kurimoto, Press-formability of stainless steel and aluminum clad sheet, J. Mater. Process. Technol. 56 (1996) 242–253.
[7] K. Yilamua, R. Hino, H. Hamasaki, F. Yoshida, Air bending and springback of stainless-steel clad aluminum sheet, J. Mater. Process. Technol. 210 (2010) 272–278.
[8] H.C. Tseng, C. Hung, C.C. Huang, An analysis of the formability of aluminum/copper clad metals with different thicknesses by the finite element method and experiment, Int. J. Adv. Manufacturing Technology 49 (2010) 1029–1036.
[9] Okui Toshiyuki, Yonemitsu Yoshihisa, Yoshida Kentarou. Development of Metal Clad Sheets and Strips. Nippon Steel & Sumitomo Metal Technical Report No. 106 July 2014.
[10] S.P. Kumar, G. Vijay, A.A. Kumar, Analytical and Numerical Prediction of Springback of SS/Al-Alloy Cladded Sheet in V-Bending, J. Manuf. Sci. Eng. 143 (2021) (2021) 031011–31021.
[11] S.P. Kumar, G. Vijay, A.A. Kumar, Experimental and numerical investigations of springback and residual stresses in bending of a three-ply clad sheet, Proc IMechE Part L: J Materials: Design and Applications. 235 (2021) (2021) 2823–2838.
[12] Hibbit, K., and Sorensen, 2007, ABAQUS/Standard Analysis User’s Manual, Hibbit, Karlsson, Sorensen Inc., Providence, USA.

IOT BASED SANITIZATION ROBOT USING RASPBERRY PI ZERO W AND UV LAMP

B Revanth Raj, Abhishek Sahani, Ravi Yadav
Mechanical Engineering with Specialization in Automobile Engineering
Delhi Technological University
Corresponded author: revanthraj3206@gmail.com

Abstract
As that's not simple to sanitise treatment room simply physical endeavour. It raises increasing risk of disease, allowing hazardous bacteria to propagate even deeper. Ultraviolet beams are being used by Ultraviolet sterilisation robotic to eliminate microbes. The live stream broadcast of the creature's environment is indeed possible. They may operate these robots through Internet, but also its Interface enables everyone to operate inside a waiting area while having to really be actually present. Most of this allows us both to sterilise the maternity ward to their specifications. Ultraviolet light kills bacteria, preventing them from multiplying by disrupting its fertility.

I. INTRODUCTION
Another purpose of green hygiene inside a surgical procedure or a client bed is really to limit germs, especially substance germs, to that of a bare max in establishing as maintaining a safe workplace for both service user and service professional.
People’s future effectively have changed dramatically as a result of something like the COVID-19 virus, but the only method to avoid this diseases from spreading further is to keep good distance as well as obey the standards set out by our individual governments. Sanitation hygiene sterilization, in particular, was becoming unrecognizable part of people's daily lives. When it comes to proper cleaning but instead sterilization, there is indeed a concern: we didn't effectively participate in the filtration even though we risk contracting a fatal disease first from polluted site. However, we could really mitigate this issue by creating a robust, productive, as well as fully independent automaton which could sterilize an area quickly often without uncovering ourself. Computers would be used in a variety of situations whereby people may be exposed. This, in just this article, we'll create an optical sterilization robot which can destroy viral coronavirus inside a facility and residential complex, but we'll use a Nintendo, several Ultraviolet light Lights, plus laser scanners to use it.

II. BACKGROUND
Currently, surgical room viruses account for 14-17% of all pathogens, whilst hospitalized diseases account for 38% of all pathogens.
As a result, regular procedural as well as postoperative scrubbing of an OR atmosphere are among the most efficient aseptic techniques for reducing the number of germs, pollen, as well as green matter in the area.

Nonetheless, because there are numerous blind spots especially unattainable regions including the interior surfaces, typical wiping routine using clean supplies for humans alone will not been capable of minimizing the quantity of these germs. A form of uv – is (uv light) has just been discovered that might help physicians in their continuing struggle to prevent bacteria in remaining in health care facilities as well as creating additional diseases.

This is a worldwide problem. Humankind has really been delayed by the influenza epidemic and quarantine. That concern of contagion from not just immediate communication but then also items believed to already be encountered by sick people had contributed to its use of anionic surfactants to disinfect packs, vessels, or even veggies. It's worth noting that ingesting dishwashing plus related compounds might cause issues for the female organism. As a result, programmed cleaning machines are needed to transport the envelopes to be sterilised on just a production line from the entering endpoint, keep those underneath Ultraviolet light for a defined amount of time, and thereafter transport out towards the output close. This mechanism will prevent washing up liquid and some other similar chemicals from accessing your female beings. The conveyor belt movement will be controlled by sensing at the entering endpoint, its storage point (within the cage), and the discharge end. During sterilization of items encased in packaging, comparable technologies will also be required in massive amount in multiple kinds of production modules, markets, and supermarkets. To just that aim, proper regulations will just be designed to incorporate conveying belts movements, Ultraviolet light bulb operation, and programmed packet/goods retention duration via device.

III. METHODOLOGY

In this chapter the block diagram of the project and design aspect of independent modules are considered. Block diagram is shown in fig.1.

Arrange all or most of the equipment on the quad rotor (UV LED, Raspberry Pi zero W, a wind turbine with batteries, Pi camera, Motor Driver) but then place your Ultraviolet rays on top of the machine. Place the Micro - controller on ahead of robotic so that the video can see why this’s next to it effectively. Later, utilizing VNC, access its window on either a faraway Computer or cell phone, and execute your Interface. The Microcontroller Board is connected to an Electric Controller in the usual sequence.

The project aims in designing a robot which is capable of monitoring and controlling from remote place. It switches on the UV lamp through which UV rays are emitted and also sanitize the surroundings. **Ultraviolet-C** Radiation, Disinfection, and Corona virus has given the current outbreak of Corona virus Disease. As there is no vaccination on the marketplace, the first and only way to survive is to take the necessary measures. It is recommended to sterilise the environment to assist slow the transmission of infection because the virus is communicated mostly by particles that arise from an affected user's nasal passage while sneeze or cough. To address these issues, we created a sanitization robot powered by a Raspberry Pi and equipped with a UV lamp. This is the internet of things (IOT) based project, where we are particularly uses the Raspberry Pi, pi camera, sanitizer ,UV lamp along with relay driver and two DC motor along with l293d motor driver with Robot chassis to build this Robotic car setup. It has a pi camera mounted over it, through which we will get live video feed and the interesting part here

is that we can control and move this robot from a web browser over the internet. The Pi cam will capture live data with regards to its surroundings and then send it to a particular IP address through internet. The user will be observing this live streaming on the mobile and according to that the user will control the robotic vehicle through the webpage and as well as control the sanitizer and UV lamp from the web page .Here relay likes as a switch to on and off the sanitizer and UV lamp.12v, 2ah battery and power bank are used to give the power supply of the robot and raspberry pi. The main controlling device of the project is raspberry pi.

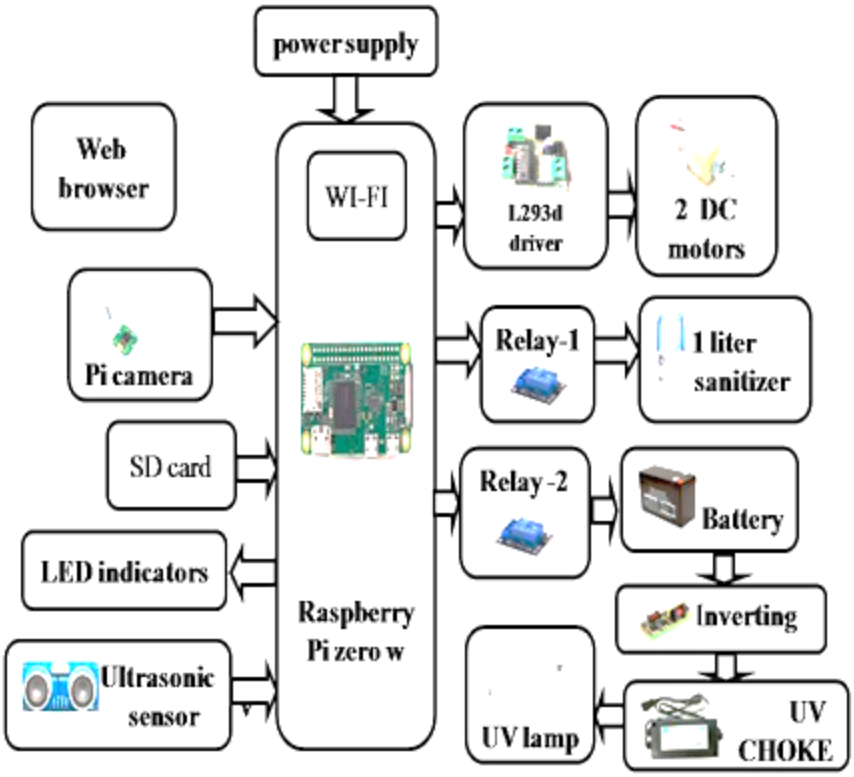


Fig 3.1: Block diagram of IOT based sanitization robot using Raspberry pi zero w and UV lamp

Hardware description

ARM processor

The Raspberry Pi Zero features a **BCM2835 chipset**, overclocked to 1 GHz with **512MB RAM**, and the same **1080p video output,** so there's plenty of oomph in that little board. It also features the same 40 pin GPIO layout as the Raspberry Pi 2/B+/A+, which you'll need to solder your own headers too. With its small form factor and reduction in connectors, the Raspberry Pi Zero **only uses ~ 140mA at 5V.**

Power supply

An AC adapter, also known as an AC/DC adapter or an AC/DC converter, is a form of external power source that is usually housed in a case that looks like an AC socket. Electrical equipment that run on electricity but doesn't have internal parts to extract the appropriate electrical power

and voltage from main grid are connected to AC adapters. An auxiliary power supply's internal circuitry is remarkably similar to the architecture of an internal power source.
External power sources should be used with technology and has no other power source as well as rechargeable batteries device, where the source could sometimes charge batteries as well as power the device when plugged in.
The need for an external power source permits battery-powered technology to be portable without any of the additional size of internal elements, and it eliminates the need to design electronics for utilization with a particular propulsion system.

UV LEDS

LEDS (Light Emitting Diode) that generates uv rays is a kind of infrared energy with wavelengths ranging from 10 (with just a velocity of roughly 30 PHz) to 400 nm (750 THz), which is longer than visible region. UV For high antimicrobial treatment, an uv - b spectrum with in spectrum of 100-400 nanometers is used. Radiation from the sun (200–280 nm), which itself is deemed the much more high antimicrobial wavelengths region based on its ability to inhibit the growth bacteria, is perhaps the most high antimicrobial wavelengths spectrum.

DC motor

The DC motor has two basic parts: the rotating part that is called the armature and the stationary part that includes coils of wire called the field coils. The stationary part is also called the stator. Figure shows a picture of a typical DC motor, Figure shows a picture of a DC armature, and Fig shows a picture of a typical stator. From the picture you can see the armature is made of coils of wire wrapped around the core, and the core has an extended shaft that rotates on bearings. You should also notice that the ends of each coil of wire on the armature are terminated at one end of the armature. The termination points are called the commutator, and this is where the brushes make electrical contact to bring electrical current from the stationary part to the rotating part of the machine.
The DC motor you will find in modem industrial applications operates very similarly to the simple DC motor described earlier in this chapter. Figure 12-9 shows an electrical diagram of a simple DC motor. Notice that the DC voltage is applied directly to the field winding and the brushes. The armature and the field are both shown as a coil of wire. In later diagrams, a field resistor will be added in series with the field to control the motor speed.
When voltage is applied to the motor, current begins to flow through the field coil from the negative terminal to the positive terminal. This sets up a strong magnetic field in the field winding. Current also begins to flow through the brushes into a commutator segment and then through an armature coil. The current continues to flow through the coil back to the brush that is attached to other end of the coil and returns to the DC power source. The current flowing in the armature coil sets up a strong magnetic field in the armature.

Motor driver

It's an embedded system semiconductor that's utilized in humanoid systems as well as embedding electronics to regulate motors. A motor driver is unquestionably anything that causes the machine to operate in accordance with the provided commands or impulses (high

and low). It takes its low power first from gamepad as well as uses that to operate a real engine that possesses a minimum power factor.

Pi camera

The Camera Unit can record high-definition videos along with taking still photos. It has video modes of 1080p30, 720p60, and VGA90, as well as always capturing. It connects to the Raspberry Pi's CSI port via a 15cm ribbon wire.

The camera is made up of a compact circuit board (25mm by 20mm by 9mm) that uses an adaptable ribbon connection to interface to the Raspberry Pi's Camera Serial Interface (CSI) bus socket. The camera's image detector is a fixed focus aperture with a native sensitivity of five megapixels. The camera's software allows still photographs up to 2592x1944 pixels and video resolutions of 1080p30, 720p60, and 640x480p60/90 pixels.

Desired specifications of the project

- The tank capacity of the Sanitizer that can be loaded is 150ml.
- The tube length of UV light designed is about 25cm.
- The Sanitizer Density is measured to be 0.90g/mL.
- The Speed obtained is 5meter/ Sec.

The brightness and time of the UV light is expressed as follows,

Brightness = Luminosity / (4Π * Distance^2)

Time = UV dose / Brightness

UV dose - $\mu Ws / cm^2$

Brightness – W / cm^2

Brightness = 15 * 4W / (4Π * 25^2) = 6W / 7853.98 cm^2

= 76.3 $\mu W / cm^2$

Thus the brightness produced by the desired project is also obtained.

Viscosity affecting the flow rate

The viscosity of a substance can alter due to a variety of circumstances, including temperature, humidity, and conditioning. A substance that performs properly in the lab could have a completely different viscosity when it exits an extreme temperature manufacturing process. If Carbopol is replaced with a green environmentally substitute, a Carbopol-based product that thins when it comes into touch with salt on the skin may lose this property.

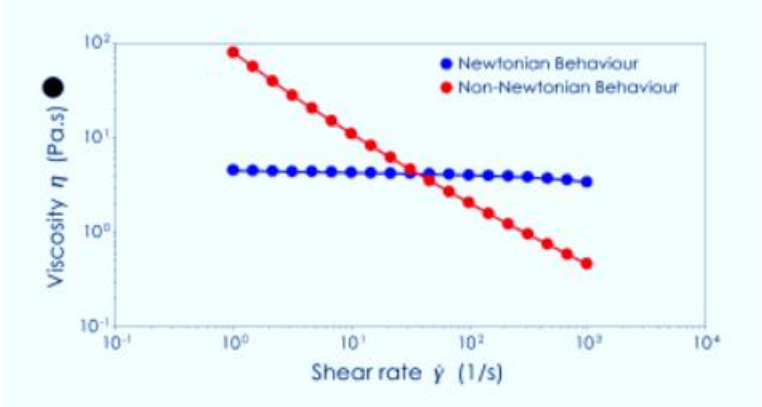


Fig.1.Viscosity of the sanitizer

The sanitizer exhibits a non - newtonian behavior, when the viscosity increases the shear rate decreases. When the viscosity increases the flow rate decreases. Thus the pressure or the force applied has to compensate to proceed with the appropriate flow rate.

IV. RESULT

This robotic can prevent injury as well as navigate about the environment unique combination of either a webcams video as well as lidar. This Ultraviolet robot has airspeed of 5 metres per second and therefore can run for 30-45 minutes on a charge without connected to with a plug socket. Hand wash Machine's Light source disinfected the majority of bacteria inside the environment, if not all. In the specified region, while there were some innocuous, quasi bacteria.

V. CONCLUSION

The above research provides a detailed assessment of robots' possibilities in biology as well as related fields, with either a focus on COVID-19 catastrophe containment. As seen in the epidemic, competent COVID-19 care can drastically limit the amount of people with the disease and casualties. Because it has now become a worldwide problem, technology driven governments may help others by giving weapon systems including automated infrastructures to ensure a successful management of the illness. Cost of health care digitalization, this analysis justifies that the adoption of medical robots had considerably improved both health of the patient’s security solutions due to manual methods. Surgical drones typically solely classified employing implementation subcategories to accommodate everything part of inpatient services, including load balanced supervision but also stable infrastructures for consistent as well as took place in order inside medical institutions.

VI. FUTURE WORK

To create the robotic smart sufficiently to prevent collisions and some other costly healthcare apparatus while efficiently disinfecting everything and every part of something like the facilities, remote sensing techniques were added to that same robotic as well as excellent charting was used with the assistance of Learning Algorithms. This will also assist in instantaneous going to charge and destination via a Photovoltaic Solar Car Charger. A autumn

technology, comprised of a camcorder and a 3D Digitizer, is included in the creature's primary functionality to recognise people who already have fallen asleep. The robot can alert health workers in the case of a serious situation involving a client recognised by the wearable sensors as well as continuous examined closely modules.

REFERENCE

1. *Gabriele Ferri, Alessandro Manzi, Pericle Salvini,Barbara Mazzolai, Cecilia Laschi, and Paolo Dario, "DustCart, an autonomous robot for door-to-door garbage collection: from DustBot project to the experimentation in the small town of Peccioli, 2011 IEEE International Conference on Robotics and Automation, Shanghai International Conference Center, May 9-13, 2011, Shanghai, China.*
2. *Hiroo Takahashi, Kojiro Iizuka, "Analysis on Weight Arrangement Scheme to Reduce theWeight of Multi-Legged Robot Takashi Kubota",Proceedings of the 2005 IEEE International Conference on Robotics and Automation, Barcelona, Spain, April 2005.*
3. *Mahmoud Tarokh and Malrey Lee,"Kinematics Modeling of Multi-legged Robots Walking on Rough Terrain" 2008 Second International Conference on Future Generation Communication and Networking Symposia.*
4. *Himadri Nath Saha, Sourav Gon, Annesha Nayak, Samabrita kundu, Sumandrita Moitra , "IoT Based Garbage Monitoring and Clearance Alert System" 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON) Pages: 204 – 208.*
5. *Bharadwaj, Alok & Yadav, Divyanshu & Varshney, Shreyshi. (2015). NON-BIODEGRADABLE WASTE – ITS IMPACT & SAFE DISPOSAL. International Journal of Advanced Technology in Engineering and Science. 3. 184- 191.*
6. *Chanprakon, P., Sae-Oung, T., Treebupachatsakul, T., Hannanta-Anan, P., & Piyawattanametha, W. (2019). An Ultra-violet sterilization robot for disinfection. 2019 5th International Conference on Engineering, Applied Sciences and Technology (ICEAST). doi:10.1109/iceast.2019.8802528*

ARTICLE IN PRESS

Ceramics International xxx (xxxx) xxx

Contents lists available at ScienceDirect

Ceramics International

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

Key role of Tb^{3+} doping on structural and photoluminescence properties of $Gd_2Ti_2O_7$ pyrochlore oxide

Umang Berwal, Vinod Singh *, Rinku Sharma

Delhi Technological University, New Delhi, 110042, India

ARTICLE INFO

Keywords:
Pyrochlore
Tb-fluorescence
Photoluminescence spectroscopy
X-ray diffraction
Raman spectroscopy

ABSTRACT

Pyrochlore structured compounds form a special class of ceramics as host matrices in phosphors with better luminescent properties. In the present work, the optical properties of pyrochlore oxide $Gd_{2-x}Tb_xTi_2O_7$ ($0 \leq x \leq 0.16$) were investigated and linked with structural properties *via* examination of structural distortion produced with the doping of rare-earth element. These oxide compositions were produced through standard solid-state reaction method. The investigations using advanced characterization techniques; X-ray diffraction and Raman spectroscopy give a signature of single phasic pyrochlore structure in the synthesized samples optimized through multiple heating protocol. The diffraction patterns and Raman spectra, both were continuously being shifted towards lower 2θ value and lower wave-number side, respectively with the doping concentration which can be interpreted by the production of defects in cationic lattice. A similar decreasing trend in particle size has been observed from W–H plot (728 Å to 616 Å) and Scanning electron micrographs (0.82 μm–0.78 μm) with the introduction of Tb^{3+} ion. The band-gap for the host matrix $Gd_2Ti_2O_7$ through UV absorption pattern came out to be 2.26 eV. From photoluminescence excitation ($\lambda_{em} = 545$ nm), a broad excitation band was observed in the wavelength range 305 nm–380 nm due to the f→d transition of 4f shell of Tb ion. Furthermore, emission spectra ($\lambda_{ex} = 375$ nm) consist of bands at 489 nm, 545 nm, 585 nm and 623 nm due to the transitions at $^5D_4 \rightarrow {}^7F_{6,5,4,}$ and 3, respectively with doping. For an optimum doping concentration with $x = 0.12$, a high intensity green emission line at 545 nm came out as $^5D_4 \rightarrow {}^7F_5$ dominating transition of terbium ion. Hence, the results explore the application of $Gd_2Ti_2O_7$ composition as Tb-fluorescence with the optimum doping concentration of 6%.

1. Introduction

A number of research groups have been widely investigating luminescent materials due to their various applications in solid-state light sources [1], computer and television screens, lasers [2–4], Hg-free lamp phosphors [5], cathode ray tube (CRT) phosphors [6,7], Plasma display panels (PDP) [8], field emission displays [9], optical thermometer in medical field [10] *etc*. For a good luminescent material, a suitable dopants drastically increases the radiative transitions over irradiative transitions in a crystalline host matrix [11]. Unlike metal ions, suitable electronic configuration of lanthanides (4f orbitals surrounded by the outer filled $5s^2$ and $5p^6$) makes them most relevant candidate as activators with surprisingly superior optical properties [12]. Therefore rare-earth based lattice structures and rare-earth ions (REI's) are the excellent candidate for both hosts and activators (luminescent centres), respectively since they show different colours of emission based on the 5d→4f and 4f→4f transitions [13]. The lanthanide elements are main contenders for being utilized as the activators with partially occupied 4f shell which is surrounded by the outer $5s^2$ and $5p^6$ orbitals. Due to this shielded $4f^N$ shell configuration of REI's, their energy levels are not much influenced by the surrounding environment irrespective of the host structure except a very slight shift in the positions and a small change in relative intensities of their spectral lines [14]. Here rare-earth based pyrochlore structured oxides were explored and investigated as host matrix to make a most efficient luminescent materials. Among ternary oxides, compound with popular formula $A_2B_2O_7$ belongs to a family of pyrochlore structure where A^{3+} and B^{4+} are cations. Rare earth pyrochlore oxides can have much importance in today's world due to their use in various fields, such as host for luminescence centres [15], catalysts [16], nuclear waste forms [17], high temperature pigments [18], ionic conductor [19], transparent ceramics [20], and water splitting [20] *etc*. Rare earth based pyrochlore oxides are cubic and ionic in nature thus their adoptability towards the variation of chemical substitution at A and B cationic position is very high. Pyrochlore oxides

* Corresponding author.
E-mail address: vinodsingh@dtu.ac.in (V. Singh).

https://doi.org/10.1016/j.ceramint.2022.04.231
Received 26 February 2022; Received in revised form 18 April 2022; Accepted 20 April 2022
Available online 25 April 2022
0272-8842/© 2022 Elsevier Ltd and Techna Group S.r.l. All rights reserved.

have various physical properties due to the fact that cation B is a transition metal atom with varying oxidation state and cation A is a REI with inert lone pair of electrons. The stability strength of pyrochlore oxide structure is described by the radii ratio of cation A (r_A) and cation B (r_B) (i.e., r_A/r_B) ratio. The ionic radii ratio for $Gd_2Zr_2O_7$ is 1.46 and for $Gd_2Ti_2O_7$ its value is 1.75. The ionic radii ratio value ranges from 1.46 to 1.78 for stable pyrochlore oxide structure [21–23]. If the value of ionic radii ratio is less than 1.46 then the structure is changed into defect fluorite structure due to the easy swapping of two cations [24]. Several pyrochlore compositions, like, $Gd_2Ti_2O_7$, $Gd_2Zr_2O_7$, $Hf_2Ti_2O_7$, $Ln_2Ti_2O_7$, *etc.* were reported with an enhanced optical property by doping with suitable dopants like Tb, Eu, V, Sm, *etc.* [25–28]. Among different pyrochlore compositions, rare earth titanates were used as phosphor materials due to easy incorporation of trivalent rare earth ions as activator at the A-site (due to nearly same ionic radii), strong luminescent, optical, magnetic, and electrical properties [29–33]. Some research groups were investigated the enhancement of optical properties of $Gd_2Ti_2O_7$ by doping of rare earth ions; namely Eu^{3+}, Er^{3+}, and Dy^{3+} at A-site [26,34–38]. Garbout et al., were reported the enhancement of photoluminescence (PL) properties and explained this phenomenon by introduction of defects due to increasing $r(REI^{3+})/r(Ti^{4+})$ in $Gd_{2-x}Eu_xTi_2O_7$ oxides with lattice expansion on replacing the Gd^{3+} (1.053 Å) by Eu^{3+} (1.066 Å). Although the compositions with more than 50% Eu^{3+} showed the luminescence quenching [26]. Pang et al., were also observed concentration quenching in thin film of Eu^{3+} doped $Gd_2Ti_2O_7$ at more than 9 at. % Eu^{3+} [36]. Furthermore, Zeljka Antic et al. have reported the improvement in luminescent spectrum as a function of annealing temperature in Eu^{3+} doped $Gd_2Ti_2O_7$ thin film [34]. Ting et al., observed increasing PL intensity in Er^{3+} ions doped $Gd_2Ti_2O_7$ host lattice up to doping concentration of 10 mol% and after that quenching effect was found [35]. Thus, these groups were reported the rare earth doped $Gd_2Ti_2O_7$ as best host material for phosphors. Therefore, optical properties of $Gd_2Ti_2O_7$ can be tuned by introducing defect fluorite structure. Terbium ions (Tb^{3+}) are used as the most supreme green light emitting luminescent activators (dopant) for various phosphors since it has an intense emission around wavelength of 544 nm with very high intensity, which is adjacent to the theoretical optimal wavelength for the green section of three primary colour centres [39]. Since the cationic radii for Gd^{3+} and Tb^{3+} are 1.053 Å and 1.040 Å, respectively. Therefore r_{Gd}/r_{Ti} and r_{Tb}/r_{Ti} are 1.740 and 1.719 respectively which shows that the doping of Tb^{3+} ions in $Gd_2Ti_2O_7$ host matrix does not affect basic pyrochlore structure. Therefore, Tb^{3+} doped $Gd_2Ti_2O_7$ composition was studied to explore their luminescent properties. Here in the present research work, we prepared pellets by the doping of Tb^{3+} rare earth trivalent ion as activator in the host $Gd_2Ti_2O_7$ lattice, using solid-state reaction method. The pellets of pristine $Gd_2Ti_2O_7$ and doped with 2%, 4%, 6%, and 8% of Tb^{3+} were studied by using characterization techniques like, X-ray diffraction spectroscopy (XRD), Micro-Raman analysis and Field Emission Scanning electron microscopy (FESEM), UV–Visible spectroscopy, and PL spectroscopy.

2. Experimental section

2.1. Sample preparation

The Tb^{3+} doped $Gd_{2-x}Tb_xTi_2O_7$ pyrochlore ternary oxides with x = 0%, 2%, 4%, 6%, and 8% were obtained by using a facile solid-state reaction route. In this method stoichiometric amount of high purity (Sigma Aldrich ~ 99.99% pure) AR-grade raw powders Gd_2O_3, TiO_2 and Tb_2O_3 were preheated for 3 h to remove moisture and extra volatile impurities. Stoichiometric number of raw powders are given in Table 1. After preheating all the powders were stoichiometrically weighed and ground intimately to obtain homogeneity using mortar and pestle. Powdered mixture of powders was pelletized using hydraulic press and dye set by applying a pressure of 4.5 Torr.

Cylindrical pellets were heated initially at a temperature of 1200 °C for 24 h duration using high temperature box furnace. The procedure of pelletizing and grinding was repeated again and again unless we get desired pure phasic structure of pyrochlore ternary oxide. For second heating protocol, all the pellets were re-grounded and pelletized again and heated at temperature 1250 °C for 36 h. To remove defects and to achieve pellets of high density, all the pellets were finally sintered at a high temperature of 1300 °C for 48 h of duration using box furnace with heating rate 5 °C/min and colling rate 2 °C/min respectively. A flow chart for the synthesis of various doping concentration of Tb^{3+} doped $Gd_2Ti_2O_7$ pyrochlore host matrix is shown in Fig. 1.

Table 1
Concentration of salts used in synthesis of $Gd_2Ti_2O_7$ doped with Tb^{3+} series.

Doping level "x"	$Gd_{2-x}Tb_xTi_2O_7$	Gd_2O_3 (gram)	TiO_2 (gram)	Tb_2O_3 (gram)
0	$Gd_2Ti_2O_7$	1.3947	0.6146	0
0.04	$Gd_{1.96}Tb_{0.04}Ti_2O_7$	1.3588	0.6110	0.028
0.08	$Gd_{1.92}Tb_{0.08}Ti_2O_7$	1.3324	0.6116	0.056
0.12	$Gd_{1.88}Tb_{0.12}Ti_2O_7$	1.3045	0.6115	0.084
0.16	$Gd_{1.84}Tb_{0.16}Ti_2O_7$	1.2766	0.6114	0.112

2.2. Characterization of Tb^{3+} doped $Gd_2Ti_2O_7$ pyrochlore oxide

The structural investigation of all the finally sintered pellets with different doping concentration of terbium were examined using X-ray diffraction technique. The XRD patterns of all the pellets were recorded on XRD machine of Bruker company (Model-D8-advance) in the range of angle 2θ from 10° to 75° using a step of 0.02° and scan speed 0.5° min^{-1} with Cu diode K_α radiation of wavelength λ = 1.5406 Å at DTU, New Delhi. All the finally sintered pellets were examined through Field emission scanning electron microscopy (FESEM-MIRA-II, TESCAN), at IUAC, New Delhi. Raman spectra for all the pellets were recorded using a laser of excitation wavelength 532 nm using the WITec Alpha 300RA Raman microscopy equipped with a 50 × objective lens, at JNU, New Delhi. Absorption spectra for pristine $Gd_2Ti_2O_7$ was examined on an UV–Visible spectrophotometer of Jasco company (Model V-770) at DTU, New Delhi. PL excitation patterns and emission spectrum for all the samples with various doping concentration of terbium were observed with a Jasco-8300 fluorescence spectrometer at room temperature using Xenon lamp.

3. Results and discussion

3.1. Structural analysis through XRD

X-ray diffraction for all the powdered samples $Gd_{2-x}Tb_xTi_2O_7$ with varying the value of x = 0.04, 0.08, 0.12, and 0.16, annealed at 1300 °C for 48 h were recorded and depicted in Fig. 2(a). A series of XRD patterns for all the samples were recorded in the 2θ range from 10° to 75° to obtain all pyrochlore superstructural peaks. The most recognizable XRD patterns containing two types of diffraction maxima. For first type of XRD peaks having high intensity peaks with miller indices (222), (400), (440), (622), (444), and (800) centered at 2θ values 30.48°, 35.34°, 50.78°, 60.35°, 63.42°, and 71.30°, respectively were represents basic fluorite structure on which pyrochlore structure was build. Furthermore, the second kind of highly ordered super-structure of pyrochlore oxide with low intensity peaks were found at 2θ values 15.15°, 29.16°, 38.62°, 46.41°, and 53.28° having miller indices planes (111), (311), (331), (511), and (531), respectively, represented by star (*) mark shown in Fig. 2. All the samples belong to cubic crystal structure with Fd-3m space group. For all the different compositions of dopant (Tb^{3+}), crystalline single phase cubic structure of pyrochlore were formed without any second phase which are in good agreement with the published results [40]. Therefore, doping of Tb^{3+} at the A-site into host $Gd_2Ti_2O_7$ does not distorted the crystal structure which indicated that the trivalent rare

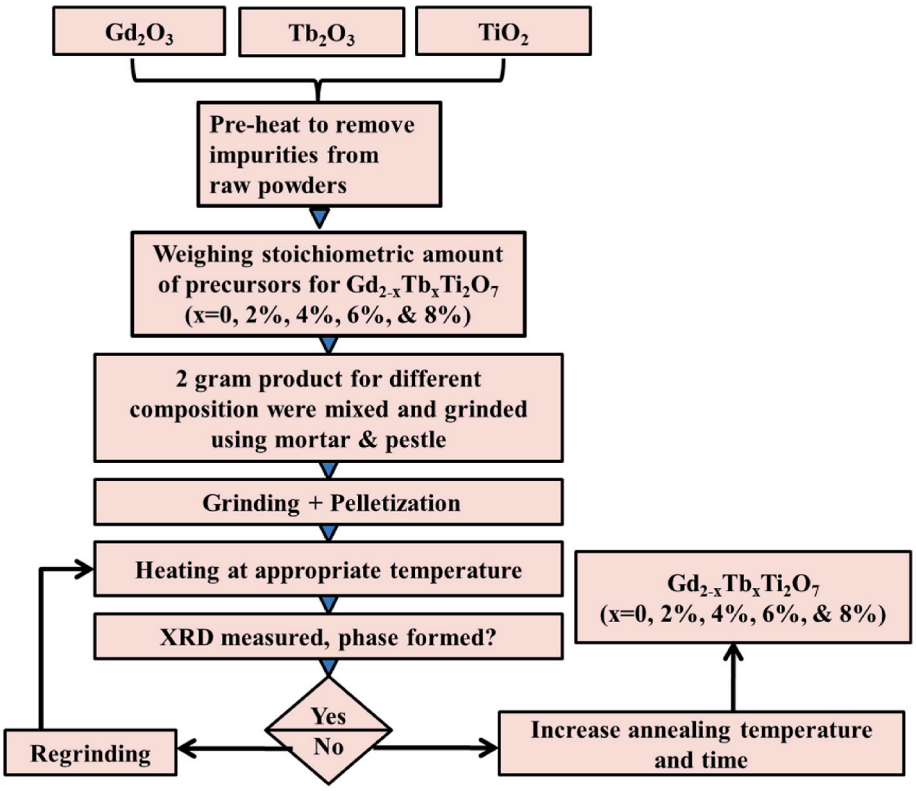


Fig. 1. Flow Chart for the synthesis of Tb^{3+} doped $Gd_2Ti_2O_7$ Pyrochlore host matrix.

earth ion Tb^{3+} have been correctly accumulated within the host matrix $Gd_2Ti_2O_7$ in place of Gd^{3+} lattice position [41]. This perfect pyrochlore structure for the synthesized solid-solution can be explained on the basis of ratio of cationic radii r_A/r_B, which can be calculated using the following relation,

$$\left[\frac{r_{Gd,Tb}}{r_{Ti}}\right] = \left[\frac{r_{Gd} \times atomic\ \% + r_{Tb} \times atomic\ \%}{r_{Ti}}\right] \quad (1)$$

where $r_{Gd} = 1.0530$ Å, $r_{Tb} = 1.0400$ Å and $r_{Ti} = 0.6050$ Å as the ionic radius for the trivalent and tetravalent cations with considering their coordination number. The calculated values of ratio of ionic radii are written in Table 2 which decreases from 1.7405 to 1.7388 on increasing Tb^{3+} concentration. An enlarged view of diffraction patterns (peak centered around 30.48° angle) within 2θ range 30° to 31° was separately plotted in Fig. 2(b) to show the shifting in diffraction patters as a result of doping. The most dominant diffraction peak (222) was slightly shifted towards higher angle side with increment of the Tb^{3+} trivalent rare earth ion in the $Gd_2Ti_2O_7$ host pyrochlore matrix as guided by a dotted line from bottom to top in Fig. 2(b). This peak shifting shows the decreasing lattice constant or the lattice shrinkage [42]. Since the cationic radius of Tb (1.040 Å) is smaller than the cationic radius of Gd (1.053 Å), hence the lattice parameter slightly decreases from 10.1508 Å to 10.1451 Å, on the substitution of Tb^{3+} at place of Gd^{3+}. XRD results concludes no phase transformation or any phase segregation in $Gd_{2-x}Tb_xTi_2O_7$ pyrochlore compositions.

From the XRD patterns, various parameters are calculated like, cationic radii ratio 'r_A/r_B', lattice parameter 'a', crystallite size 'D', and strain 'ε' and tabulated in Table 2. The average crystallite size and strain can be calculated using following Williamson Hall (WH) relation;

$$\beta_{hkl} \cos\theta = \frac{0.94*\lambda}{D} + 4\varepsilon \sin\theta \quad (2)$$

here 'β_{hkl}', 'ε', and 'λ' represents the Full Width Half Maximum (FWHM) of the first five dominating XRD peaks corresponding to their θ values, micro lattice strain related to the synthesized pyrochlore oxide samples, and wavelength of the X-rays used in XRD machine ~1.5046 Å, respectively. The factor 0.94 represents the value of structure factor for spherical crystals having cubic symmetry and 'D' represents crystallite size.

To find the crystallite size and micro-strain for pristine and Tb^{3+} doped pyrochlore host matrix $Gd_2Ti_2O_7$, a straight-line graph is plotted between $\beta_{hkl}\cos\theta$ on y-axis and $\sin\theta$ values on X-axis as shown in Fig. 2 (c) for all the compositions. The values of crystallite size 'D' and micro-strain 'ε' can be calculated by the y-intercept and slope values respectively from straight line of WH plot [43,44].

The crystallite size and strain were calculated from the observed values of intercept and slope, respectively. It was observed that the value of crystallite size decreases from 727.7 ± 4.3 Å to 616.2 ± 4.3 Å. Furthermore, it was observed that with increasing Tb^{3+} content in $Gd_2Ti_2O_7$ host matrix, the value of strain continuously decreases as shown by Fig. 2(d). This decreasing grain size of the Tb-doped composition originates from the doping of smaller cation, which relaxes the cationic lattice. Thus, both crystallite size and strain decrease upon doping, although doping concentration is significantly less; therefore, both parameters vary a very small.

3.2. FESEM analysis

The FESEM micrographs of pristine samples of $Gd_{2-x}Tb_xTi_2O_7$ (x = 0.04, 0.08, 0.12, and 0.16) annealed at 1300 °C were depicted in Fig. 3 (a–e). Grain size for all the synthesized compositions were studied under high magnification of the order of 20,000 times. These FESEM micrographs were analyzed through ImageJ software. All FESEM images show grain size distribution in the range of 400 nm–1400 nm. It was observed that all the samples have spherical agglomerated particles with almost equal free space. Although all samples have shown very compact morphology at the surface. The exact grain size was found to be 0.82 ± 0.04 μm, 0.81 ± 0.01 μm, 0.82 ± 0.02 μm, 0.79 ± 0.01 μm, and 0.78 ±

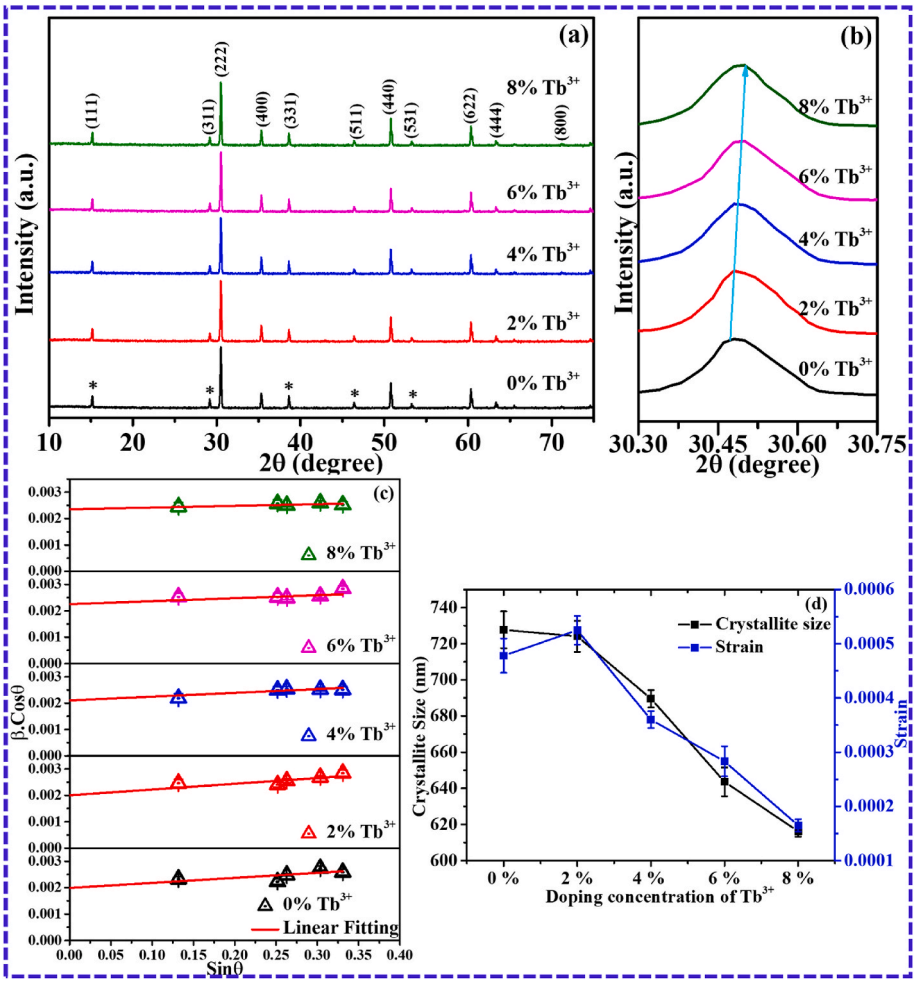


Fig. 2. (a). A set of XRD spectra for undoped and Tb^{3+} doped samples of $Gd_2Ti_2O_7$, (b). Peak shifting in (222) peak, (c). WH plot and (d). Variation in the values of crystallite size and strain with doping concentration of Terbium.

Table 2
Represents the structural parameters for undoped and doped samples of $Gd_2Ti_2O_7$ calculated from XRD results.

Doping level "x"	$Gd_{2-x}Tb_xTi_2O_7$	r_A/r_B	Intercept	Slope	Crystallite Size in Å (D)	Strain
0	$Gd_2Ti_2O_7$	1.7405	0.00199 ±0.00033	0.00191 ±0.00012	727.7 ± 4.3	0.00048 ± 0.00003
0.04	$Gd_{1.96}Tb_{0.04}Ti_2O_7$	1.7401	0.00200 ±0.00030	0.00210 ±0.00011	724.1 ± 4.0	0.00053 ± 0.00002
0.08	$Gd_{1.92}Tb_{0.08}Ti_2O_7$	1.7396	0.00210 ±0.00018	0.00144 ±0.00026	689.6 ± 3.2	0.00036 ± 0.00004
0.12	$Gd_{1.88}Tb_{0.12}Ti_2O_7$	1.7392	0.00225 ±0.00031	0.00113 ±0.00011	643.6 ± 2.7	0.00028 ± 0.00001
0.16	$Gd_{1.84}Tb_{0.16}Ti_2O_7$	1.7388	0.00235 ±0.00012	0.00066 ±0.00005	616.2 ± 2.6	0.00016 ± 0.00001

0.02 μm for the samples $Gd_2Ti_2O_7$, $Gd_{1.96}Tb_{0.04}Ti_2O_7$, $Gd_{1.92}Tb_{0.08}Ti_2O_7$, $Gd_{1.88}Tb_{0.12}Ti_2O_7$ and $Gd_{1.84}Tb_{0.16}Ti_2O_7$, respectively as shown in Fig. 3(f–j). Thus, it was found that all the samples have almost equal grain size. The micrographs corresponding to all composition show almost similar void volume which provides an indirect information about nearly same bulk density. The estimated equal grain size and similar compact morphology convey can be validated by same heating protocol for the whole series. Although, on comparing the micrographs of pure host and 8% Tb^{3+} doped host, the doping has introduced some structural distortion which did not induce any significant change in grain size and surface morphology as studied in FESEM analysis of the series $Gd_{2-x}Tb_xTi_2O_7$ (x = 0.04, 0.08, 0.12, and 0.16). Furthermore, all the elements (Gd, Ti, Tb, and O) are uniformly distributed as represented by the elemental mapping through energy dispersive x-ray analysis (EDX) and shown in Fig. 3(k–n) for the last composition $Gd_{1.84}Tb_{0.16}Ti_2O_7$ (x = 0.16). Thus, EDX analysis concludes that the all compositions have uniform distribution of the dopant (Tb^{3+}) throughout the host lattice $Gd_2Ti_2O_7$.

3.3. Raman spectroscopic analysis

For micro-structural analysis of vibrations of Tb^{3+} doped $Gd_2Ti_2O_7$ samples, Raman spectra were collected at room temperature in the wavenumber range of 100–1000 cm^{-1} for $Gd_2Ti_2O_7$ host and various concentrations of Tb^{3+} doped $Gd_2Ti_2O_7$ samples. Fig. 4 (a) shows a series of Raman spectra for undoped and doped compositions and

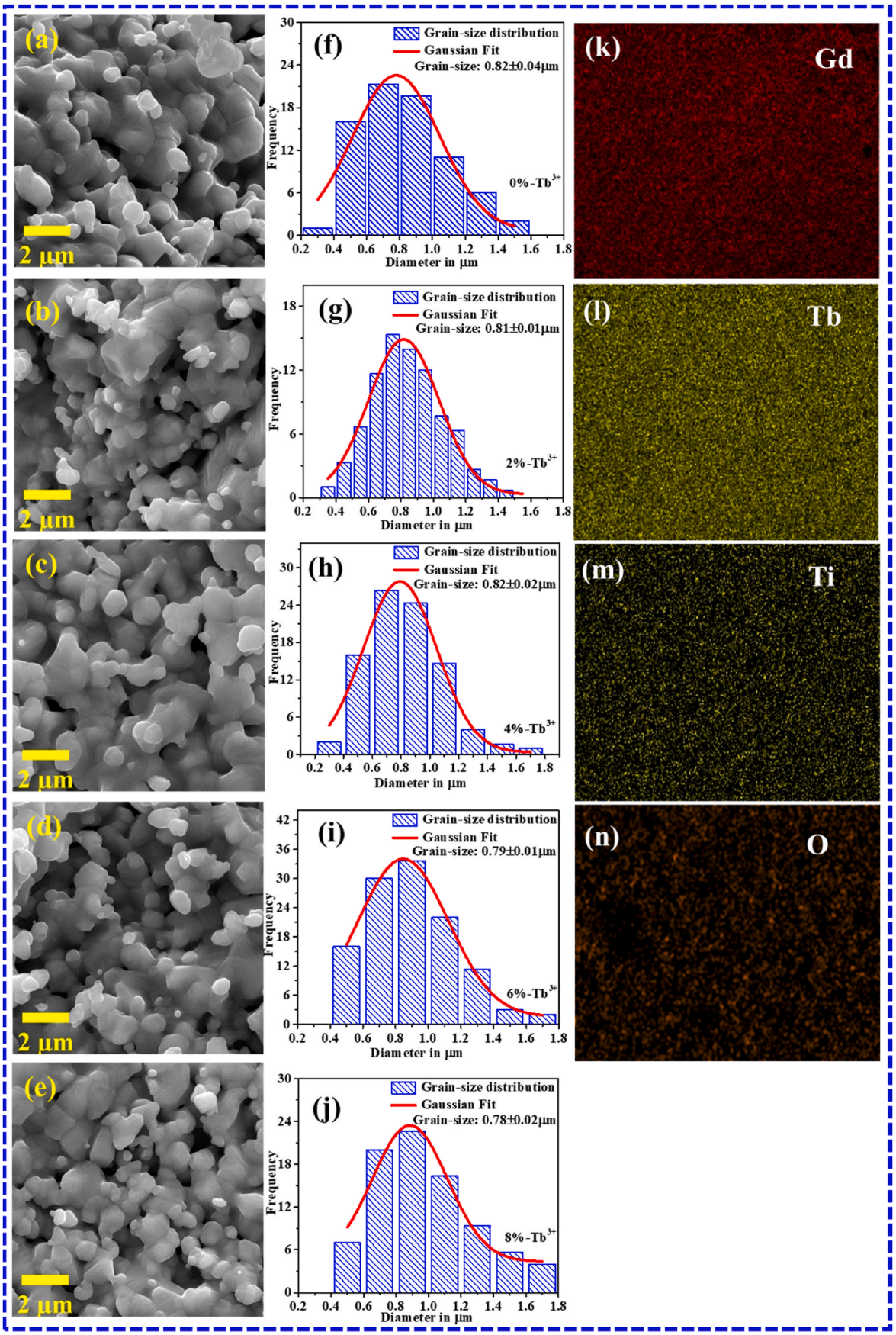


Fig. 3. (a–e). FESEM micrographs and (f–j). Corresponding grain size histogram for the samples of $Gd_{2-x}Tb_xTi_2O_7$ (x = 0.04, 0.08, 0.12, and 0.16). The elemental mapping is shown in (k–n) for the last composition $Gd_{1.84}Tb_{0.16}Ti_2O_7$ (x = 0.16).

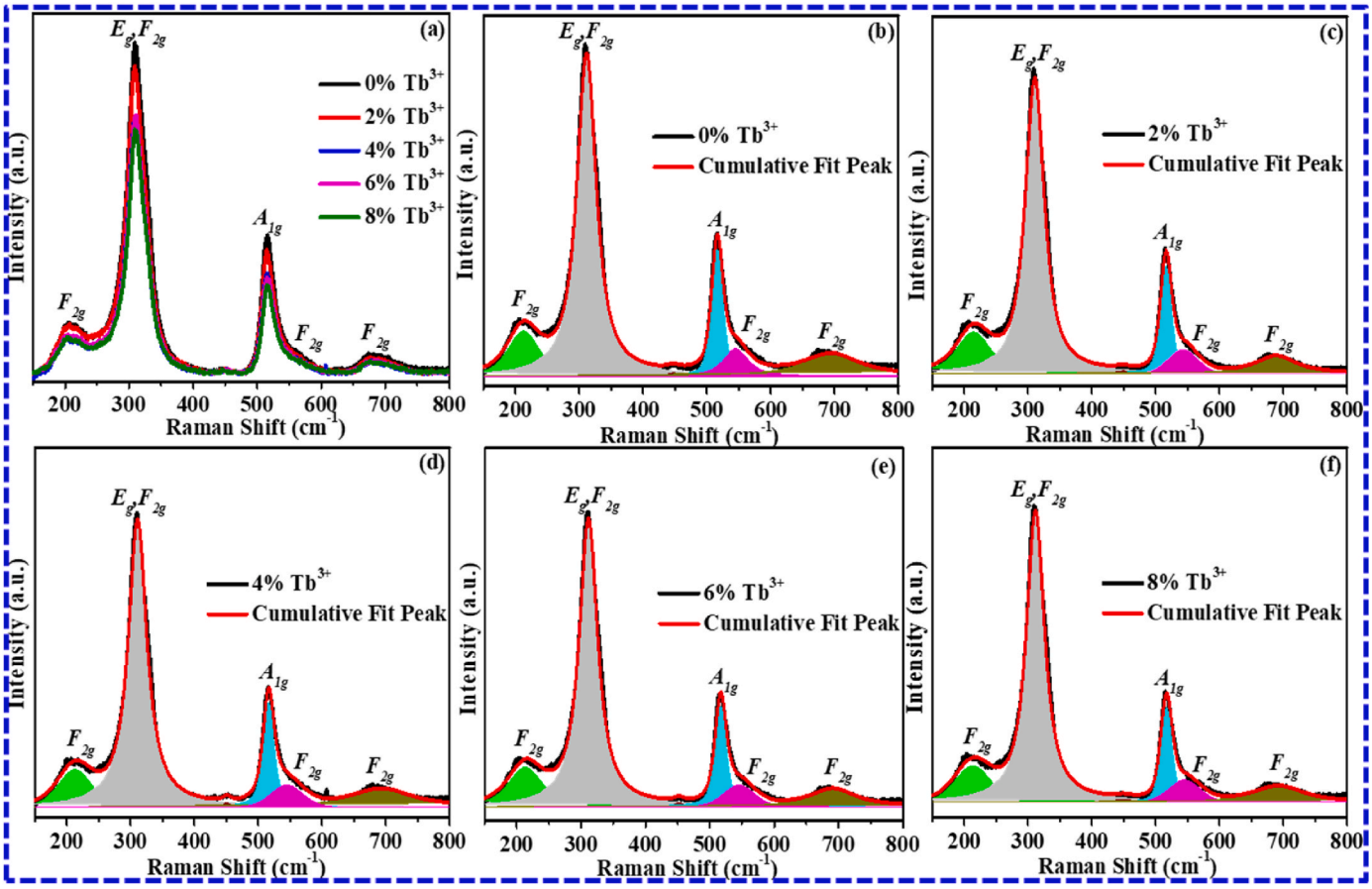


Fig. 4. (a). A set of Raman spectra for undoped and Tb^{3+} doped samples of $Gd_2Ti_2O_7$ and (b–f). Cumulative peak fitting for all compositions.

Figures (b, c, d, e, and f) show the deconvoluted spectrum into various vibrational active modes of pyrochlore structure. The Raman spectra of all the various concentration of Tb^{3+} doped $Gd_2Ti_2O_7$ having the same signature peaks as that of undoped $Gd_2Ti_2O_7$ pyrochlore host. All the Raman spectra shows that Tb^{3+} were completely dissolved in $Gd_2Ti_2O_7$ host without too much structural distortion which is in good agreement with the XRD outcomes. General formula, $A_2B_2O_7$, pyrochlores with Fd-3m space group have generally six Raman active modes and seven infrared active modes [45]. The pyrochlore structure with Fd-3m space group have only six Raman active modes at the brillion zone center: $\Gamma = A_{1g} + E_g + 4T_{2g}$ [46]. These all modes of vibrations are due to motion of oxygen atom only because cation A and B motion give rise to active vibrational Raman modes.

Five vibrational active modes ($A_{1g} + E_g + 3T_{2g}$) are due to the oxygen vibrations centered at 48f site of 'O', while the sixth vibrational mode T_{2g} is due to the oxygen vibration centered at 8b site of 'O' [47]. The most intense band at frequencies around (209–213) cm^{-1}, (310–312) cm^{-1}, (516–517) cm^{-1}, are due to the F_{2g}, E_g and A_{1g} modes, respectively, and all these three most intense bands as shown in Fig. 4 have good agreement with the published studies [48]. The remaining three Raman peaks of F_{2g} modes are difficult to identify due to their small intensity. The most intense band around frequency 311 cm^{-1} is due the bending of O-Gd-O mode and results from the contribution of two modes *i.e.,* E_g and F_{2g} modes and the peak around 516 cm^{-1} is due to the internal stretching of Gd–O mode of A_{1g} vibrations. Less intense peaks observed at wavenumber nearly 212 cm^{-1} and 545 cm^{-1} are due to the remaining F_{2g} modes of vibrations. A very small intensity peak around wavenumber 450 cm^{-1} is due to the trace amounts of TiO_2 [49]. Peak around wavenumber 685 cm^{-1} are due to the Ti–O stretching and distortion of TiO_6 octahedra [48]. As we observed from the XRD data that the lattice constant monotonically decreases from 10.1508 Å to 10.1451 Å as a function of Tb-doping with decreasing r_A/r_B. This behaviour can be corelated with force constant between the various Raman bands. Decreasing lattice constant means increasing force constant which implies that the value of bond length decreases which is validated by blue shift in Raman active modes. The estimated peak shifting as a function of Tb-doping is presented in Table 3.

3.4. UV–visible spectroscopic analysis

UV–Visible absorption pattern for undoped $Gd_2Ti_2O_7$ host pyrochlore oxide sample was recorded within the wavelength range 200–700 nm and depicted in Fig. 5 (a). It is the noble technique for the

Table 3
Represents the structural parameters for undoped and doped samples of $Gd_2Ti_2O_7$ calculated from XRD results.

Doping	Tb^{3+} doped $Gd_2Ti_2O_7$ (Peak Shifting $\Delta\omega$ cm^{-1})					
	1st Peak	2nd Peak	3rd Peak	4th Peak	5th Peak	6th Peak
0%	0.0	0.0	0.0	0.0	0.0	0.0
2%	0.3_B	0.7_B	3.2_R	0.1_R	5.8_R	2.8_B
4%	1.5_B	1.1_B	1.9_R	0.4_B	4.8_B	3.4_B
6%	1.9_B	0.4_B	6.2_B	0.5_B	0.5_R	7.8_B
8%	2.5_B	1.0_B	3.7_B	0.2_B	0.5_R	2.7_B

The most intense intensity peaks around 311 cm^{-1} and 516 cm^{-1} slightly shifts towards higher wavenumber side. Thus, a blue shift is observed from the deconvolution of Raman spectra as on doping of Tb^{3+} to $Gd_2Ti_2O_7$. The bond length decrement produced by doping terbium at the place of gadolinium can also be interpreted by the production of cation anti-site defects. Like XRD results, Raman spectroscopic analysis of Tb^{3+} doped $Gd_2Ti_2O_7$ samples concludes that the pyrochlore structure is present in all synthesized samples without showing any structural departure on the doping of terbium ion except some peak shifting.

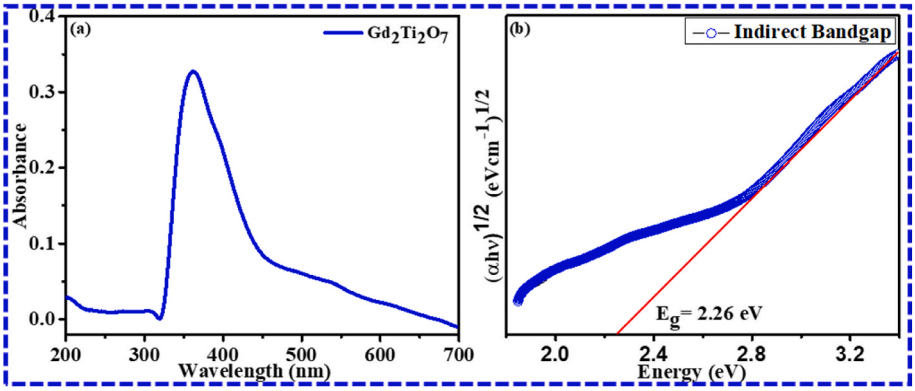


Fig. 5. UV–Visible Absorption pattern and tauc-plot for the undoped sample $Gd_2Ti_2O_7$.

determination of band-gap. It is clearly visible from the UV spectrum that the powder of $Gd_2Ti_2O_7$ shows around 32% optical absorbance. A sharp high intensity absorption peak for undoped sample ($Gd_2Ti_2O_7$) was observed in the UV region at 362 nm due to intrinsic transitions of host matrix $Gd_2Ti_2O_7$. This optical absorbance is stretched in the UV region around 325–425 nm. The band-gap can be determined by extraploting the linear region of the graph drawn between $(\alpha h\nu)^{1/n}$ vs $h\nu$. The obtained expression is known as Tauc plot can be explained by the following relation [50],

$$(\alpha h\nu)^{1/n} = k\left(h\nu - E_g\right) \tag{3}$$

here 'α' is the absorption coefficient; 'hν' is the photon incident energy, 'n' is nature of the transition, 'E_g' is the band gap energy, and 'K' is the energy independent constant. Value of n = 1/2 for direct allowed transitions and n = 2 for indirect allowed transitions. Pyrochlore oxides have indirect band gap so we select n = 2 and the Tauc plot for the undoped sample of $Gd_2Ti_2O_7$ is shown in Fig. 5 (b).

Therefore, using Tauc plot, band-gap is calculated by extrapolating linear portion of the $(\alpha h\nu)^{1/2}$ versus hν curve and the extrapolating line hits on the energy axis at the value of 2.26 eV. Hence, the prepared sample of $Gd_2Ti_2O_7$ has band-gap of 2.26 eV.

3.5. PL spectroscopic analysis

Terbium ion is a popular activator which provides intense photo-luminescence in the 450–650 nm region for a large number of ceramic compounds. Generally, terbium is stable trivalent ion (Tb^{3+}) state, used as an activator of green emission as it has a sharp high intensity emission peak around 545 nm. Luminescence phenomenon of a material is monitored by using PL results. PL properties were illustrated in Fig. 6 for host $Gd_2Ti_2O_7$ and in Fig. 7 for Tb^{3+} doped $Gd_2Ti_2O_7$. An excitation band was recorded by selecting the emission of Tb^{3+} transition ($^5D_4 \rightarrow {}^7F_5$) at a wavelength of 545 nm for the host $Gd_2Ti_2O_7$ (shown in Fig. 6 a) and various Tb^{3+} doped $Gd_2Ti_2O_7$ compositions (shown in Fig. 7 a). From the excitation result, a broad band is observed in the wavelength range from 305 nm to 380 nm peaking at ~375 nm that can be attributed to the f → d transitions of 4f shell of Tb^{3+} in the excitation, which is in good agreement with the previous results [15]. An improvement in the excitation intensity peak is observed with the terbium ion doping in the host lattice of $Gd_2Ti_2O_7$. The Gd–O charge transfer band and the general f-d transition lines of the terbium ion at the longer wavelength region have not been observed due to their relatively week intensity compared to strong $Gd_2Ti_2O_7$ host excitation band in undoped samples.

Furthermore, emission spectrum of terbium doped and undoped samples of $Gd_2Ti_2O_7$ was recorded at the excitation wavelength of 375 nm. An emission peak around 439 nm is observed in the emission band of the host $Gd_2Ti_2O_7$ due to the presence of Gd^{3+} ion at cationic lattice in pyrochlore oxide. Emission spectra of Tb^{3+} doped $Gd_2Ti_2O_7$ samples were monitored in the wavelength range 395 nm–700 nm by using excitation wavelength (λ_{ex} = 375 nm), shown in Fig. 7 b. It can be noted from the Pl emission bands of Tb^{3+} doped compositions consist of the dominant transitions at 489 nm, 545 nm, 585 nm, and 623 nm.

A schematic diagram for optical energy transitions of terbium ion is represented in Fig. 8. Trivalent ion Gd^{3+} have no tendency to give out luminescence spectrum in visible region. When Tb^{3+} ion doped into the $Gd_2Ti_2O_7$ host lattice and excited to low wavelength, the energy is absorbed by the host lattice and deliver it to Tb^{3+} ions. With increasing Tb^{3+} doping, at the low wavelength firstly intensity peak at 439 nm increases until all the energy being transferred. Excessive doping of Tb^{3+} ions into the host lattice doesn't increase the intensity and the host lattice $Gd_2Ti_2O_7$ reduces the absorbed energy by the process of charge transfer, hence the peak intensity at the low wavelength 439 nm

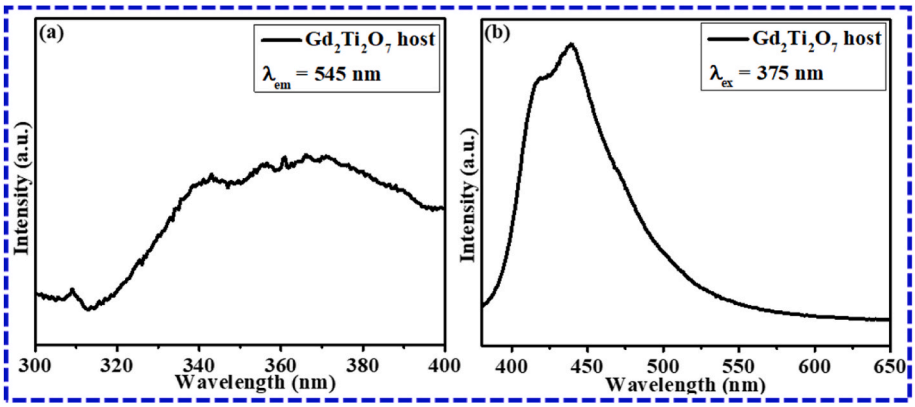


Fig. 6. (a) Excitation and (b) Emission patterns for host $Gd_2Ti_2O_7$.

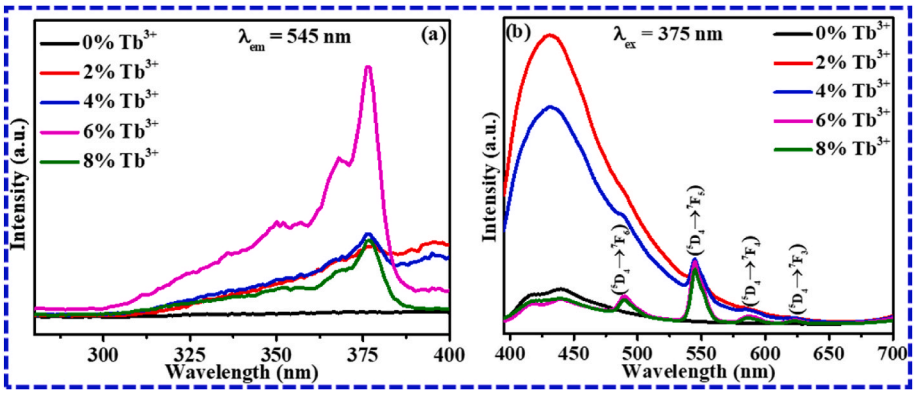


Fig. 7. (a) Excitation (λ_{em} = 545 nm) and (b) Emission (λ_{ex} = 345 nm) pattern of $Gd_2Ti_2O_7$ doped Tb^{3+} for various concentrations.

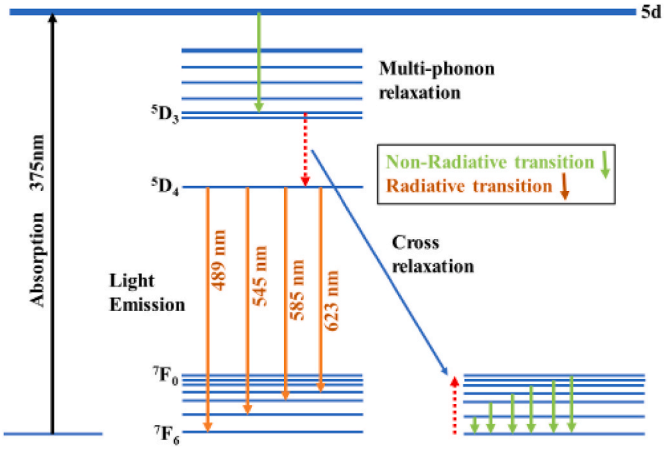


Fig. 8. Tb_2O_3 Energy level diagram.

decreases over the maximum saturation limit of 4% Tb^{3+} doping. Further for 6% terbium doping, an electron under excitation of 375 nm goes into the 5d localised band first and returns back to $^5D_3$ energy level non-radiatively [51]. Due to high concentration of Tb^{3+}, $^5D_3$ to $^5D_4$ transition is resonant with the $^7F_6$ to $^7F_{0,1}$ transition. The energy from the $^5D_3 \rightarrow {}^7D_4$ transition may be taken up by a neighbouring ion *via* a cross relaxation process [52]. As the doping concentration of terbium ion at the A-site is further increased, then peaks at 489 nm ($^5D_4 \rightarrow {}^7F_6$), 545 nm ($^5D_4 \rightarrow {}^7F_5$), 585 nm ($^5D_4 \rightarrow {}^7F_4$), and 623 nm ($^5D_4 \rightarrow {}^7F_3$) are appeared due to the transition from $^5D_4 \rightarrow {}^7F_{6,5,4, \text{ and } 3}$ trivalent Tb^{3+} in the $Gd_2Ti_2O_7$ pyrochlore phosphor. The emission line $^5D_4 \rightarrow {}^7F_5$ (high intensity green line), at 545 nm, is dominating over the other emission lines. This is due to the fact that for even higher doping concentration of trivalent terbium ion in the host $Gd_2Ti_2O_7$, there is absence of blue emission since cross relaxation give rise to an enhancement in the population of the $^5D_4$ state at the loss of $^5D_3$ state [51]. The emitting green light intensity increases with doping of Tb^{3+} in $Gd_2Ti_2O_7$ host up to 6%, after that again decreasing with further doping of Tb^{3+} due to the concentration quenching effect [28]. The distance between the terbium ions decreases when the ion concentration increased up to (6% Tb^{3+}) a maximum limit. As the distance between the Tb^{3+} ions decrease, energy transfer between the two Tb^{3+} ions are easy, therefore, due to this non-radiative transition results. Hence the emission intensity is decreased for 8% Tb^{3+} doping in host $Gd_2Ti_2O_7$. It is evident from PL studies that the $Gd_2Ti_2O_7$ pyrochlore oxide can be used as host matrix for Tb-fluorescence with a suitable doping concentration of 6%.

4. Conclusion

The single phasic pyrochlore structured $Gd_2Ti_2O_7$ doped Tb^{3+} samples were prepared by solid-state reaction method with the inclusion of multistage heating protocol. XRD and Raman spectra at room temperature were used for the confirmation of crystalline pyrochlore phase present in all the finally heated pellets. XRD traces revealed that samples annealed at temperature 1300 °C exhibited all the diffraction peaks pertaining to single-phase of pyrochlore with space group symmetry Fd-3m. The crystallite size changes from 727.7 nm (for $Gd_2Ti_2O_7$) to 616.2 nm (for 8% Tb doping) and the lattice parameter slightly decreases from 10.1508 Å to 10.1451 Å with doping of Tb^{3+} ions at A-site of the host matrix. FESEM micrographs of all the samples consist of agglomerated spherically shaped particles having the diameter of the order of micron. Further, Raman spectra shows that Tb^{3+} were completely dissolved in $Gd_2Ti_2O_7$ host without too much structural distortion which is in good agreement with the XRD outcomes. Also most intense Raman peaks at 311 cm^{-1} and 516 cm^{-1} slightly shifts towards higher wavenumber side (blue shift), due to the decrement of bond length Gd–O with terbium doping in host lattice. The bond length reduction is due to the production of cation anti-site defects. Optical study using UV-spectroscopy provides a value of indirect band-gap for $Gd_2Ti_2O_7$ as 2.26 eV. From PL results, emission band of the host $Gd_2Ti_2O_7$ consists of an intense transition at 439 nm attributed to Gd^{3+} ion. Moreover, the emission band of Tb^{3+} doped samples consist of four transitions 489 nm ($^5D_4 \rightarrow {}^7F_6$), 545 nm ($^5D_4 \rightarrow {}^7F_5$), 585 nm ($^5D_4 \rightarrow {}^7F_4$), and 623 nm ($^5D_4 \rightarrow {}^7F_3$) in the visible region which are related to transitions among the levels in Tb^{3+} dopant. It is concluded from the PL results that green light emission of the Tb^{3+} ion is dominating by $^5D_4 \rightarrow {}^7F_5$ transition in host matrix due to energy transfer from host to Tb^{3+} ions upon excitation at λ_{em} = 545 nm. Thus, the aforementioned PL studies explore the application of Tb doped $Gd_2Ti_2O_7$ up to a maximum concentration of 6% as green phosphor with the visible emission peaks without undergoing any phase transition or structural distortion.

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.

Acknowledgement

The author, Umang Berwal is thankful to Delhi Technological University, New Delhi for providing Junior Research Fellowship. The authors are also thankful to Delhi Technological University, New Delhi for the measurements XRD, UV spectroscopy and PL spectroscopy. FESEM facility of Inter-University Accelerator Centre is also acknowledged. The authors are thankful to Dr. Pawan Kumar Kulriya, Jawaharlal Nehru

University for scientific discussion. The authors acknowledge the time of Dr. Amit Kumar Vishwakarma, Atma Ram Sanatan Dharma College, University of Delhi and Dr. Vinay Kumar, Department of Electronics and Communication Engineering, Graphic Era (Deemed to be University) for scientific discussion regarding optical studies.

References

[1] R.-J. Xie, Y.Q. Li, N. Hirosaki, H. Yamamoto, Nitride Phosphors and Solid-State Lighting, 2016, https://doi.org/10.1201/b10939.
[2] S. Babu, V. Reddy Prasad, D. Rajesh, Y.C. Ratnakaram, Luminescence properties of Dy3+ doped different fluoro-phosphate glasses for solid state lighting applications, J. Mol. Struct. 1080 (2015) 153–161, https://doi.org/10.1016/j.molstruc.2014.09.080.
[3] Y. Fujimoto, M. Nakatsuka, Infrared luminescence from bismuth-doped silica glass, Jpn. J. Appl. Phys. 40 (2001) 14–17, https://doi.org/10.1143/jjap.40.l279.
[4] M. Gaft, G. Panczer, R. Reisfeld, E. Uspesnsky, Laser-induced time-resolved luminescence as a tool for rare-earth element identification in minerals, Phys. Chem. Miner. 28 (2001) 347–363, https://doi.org/10.1007/s002690100163.
[5] V.V. Shinde, R.G. Kunghatkar, S.J. Dhoble, UVB-emitting Gd^{3+}-activated M_2O_2S (where M = La, Y) for phototherapy lamp phosphors, Luminescence 30 (2015) 1257–1262, https://doi.org/10.1002/bio.2889.
[6] A. Vecht, C. Gibbons, D. Davies, X. Jing, P. Marsh, T. Ireland, J. Silver, A. Newport, D. Barber, Engineering phosphors for field emission displays, J. Vac. Sci. Technol., B Microelectron. Nanometer Struct. Process. Meas. Phenom. 17 (1999) 750, https://doi.org/10.1116/1.590633.
[7] L. Ozawa, M. Itoh, Cathode ray tube phosphors, Chem. Rev. 103 (2003) 3835–3855, https://doi.org/10.1021/cr0203490.
[8] L.Y. Zhou, F.Z. Gong, J.X. Shi, M.L. Gong, H. Bin Liang, A novel red phosphor $Na_2Ca_4Mg_2Si_4O_{15}:Eu^{3+}$ for plasma display panels, Mater. Res. Bull. 43 (2008) 2295–2299, https://doi.org/10.1016/j.materresbull.2007.08.014.
[9] C. Feldmann, T. Jüstel, C.R. Ronda, P.J. Schmidt, Inorganic luminescent materials: 100 Years of research and application, Adv. Funct. Mater. 13 (2003) 511–516, https://doi.org/10.1002/adfm.200301005.
[10] D. Manzani, J.F.D.S. Petruci, K. Nigoghossian, A.A. Cardoso, S.J.L. Ribeiro, A portable luminescent thermometer based on green up-conversion emission of Er^{3+}/Yb^{3+} co-doped tellurite glass, Sci. Rep. 7 (2017) 1–11, https://doi.org/10.1038/srep41596.
[11] G. Blasse, B.C. Grabmaier, A general introduction to luminescent materials, in: Lumin. Mater, Springer, 1994, pp. 1–9.
[12] G.H. Dieke, R.A. Satten, Spectra and energy levels of rare earth ions in crystals, Am. J. Phys. 38 (1970) 399–400.
[13] G. Blasse, Luminescence of rare earth ions at the end of the century, J. Alloys Compd. 192 (1993) 17–21, https://doi.org/10.1016/0925-8388(93)90172-J.
[14] G. Blasse, B.C. Grabmaier, Energy transfer, in: Lumin. Mater., Springer, 1994, pp. 91–107.
[15] J. Yang, Y. Su, J. Li, Novel tube-like $Y_2Sn_2O_7:Tb^{3+}$ crystals: hydrothermal synthesis and photoluminescence properties, Chem. Lett. 39 (2010) 182–183, https://doi.org/10.1246/cl.2010.182.
[16] J. Zeng, H. Wang, Y.C. Zhang, M.K. Zhu, H. Yan, Hydrothermal synthesis and photocatalytic properties of pyrochlore $La_2Sn_2O_7$ nanocubes, J. Phys. Chem. C 111 (2007) 11879–11887, https://doi.org/10.1021/jp0684628.
[17] A. Kumar, P.K. Kulriya, S.K. Sharma, V. Grover, A.K. Tyagi, V.K. Shukla, Structural and compositional effects on the electronic excitation induced phase transformations in $Gd_2Ti_{2-y}Zr_yO_7$ pyrochlore, J. Nucl. Mater. 539 (2020), 152278, https://doi.org/10.1016/j.jnucmat.2020.152278.
[18] F. Matteucci, G. Cruciani, M. Dondi, G. Baldi, A. Barzanti, Crystal structural and optical properties of Cr-doped $Y_2Ti_2O_7$ and $Y_2Sn_2O_7$ pyrochlores, Acta Mater. 55 (2007) 2229–2238, https://doi.org/10.1016/j.actamat.2006.11.008.
[19] M.R. Díaz-Guillén, K.J. Moreno, J.A. Díaz-Guillén, A.F. Fuentes, K.L. Ngai, J. Garcia-Barriocanal, J. Santamaria, C. Leon, Cation size effects in oxygen ion dynamics of highly disordered pyrochlore-type ionic conductors, Phys. Rev. B Condens. Matter 78 (2008) 1–7, https://doi.org/10.1103/PhysRevB.78.104304.
[20] T. Yanagida, G. Okada, Characterizations of optical properties and radiation induced luminescence of Bi-doped $La_2Zr_2O_7$ transparent ceramics, J. Ceram. Soc. Jpn. 124 (2016) 564–568, https://doi.org/10.2109/jcersj2.15237.
[21] J. Lian, K.B. Helean, B.J. Kennedy, L.M. Wang, A. Navrotsky, R.C. Ewing, Effect of Structure and Thermodynamic Stability on the Response of Lanthanide Stannate Pyrochlores to Ion Beam Irradiation, 2006, pp. 2343–2350.
[22] M.A. Subramanian, G. Aravamudan, G.V.S. Rao, Oxide Review, 1983, p. 15.
[23] J. Lian, L.M. Wang, S.X. Wang, J. Chen, L.A. Boatner, R.C. Ewing, Nanoscale manipulation of pyrochlore : new nanocomposite ionic conductors, Phys. Rev. Lett. (2001) 3–6, https://doi.org/10.1103/PhysRevLett.87.145901.
[24] K.E. Sickafus, K.E. Sickafus, L. Minervini, R.W. Grimes, J.A. Valdez, Radiation Tolerance of Complex Oxides, Science (2013) 748, https://doi.org/10.1126/science.289.5480.748.
[25] Y. Zhang, D. Linghong, P. Xinling, W. Zhang, Influence of annealing temperature on luminescent properties of Eu^{3+}/V^{5+} co-doped nanocrystalline $Gd_2Ti_2O_7$ powders, J. Rare Earths 27 (2009) 900–904.
[26] A. Garbout, M. Férid, Influence of Eu substitution for Gd on the structure and photoluminescent properties of $(Gd_{1-x}Eu_x)_2Ti_2O_7$ pyrochlore solid solutions, Adv. Powder Technol. 28 (2017) 1382–1390.
[27] P.A. Tanner, Some misconceptions concerning the electronic spectra of tri-positive europium and cerium, Chem. Soc. Rev. 42 (2013) 5090–5101, https://doi.org/10.1039/c3cs60033e.
[28] M.L. Pang, J. Lin, J. Fu, Z.Y. Cheng, Luminescent properties of $Gd_2Ti_2O_7$: Eu^{3+} phosphor films prepared by sol–gel process, Mater. Res. Bull. 39 (2004) 1607–1614.
[29] J.E. Greedan, Frustrated rare earth magnetism: spin glasses, spin liquids and spin ices in pyrochlore oxides, J. Alloys Compd. 408–412 (2006) 444–455, https://doi.org/10.1016/j.jallcom.2004.12.084.
[30] H. Matsushita, S. Nomura, S. Ando, S. Endo, T. Irie, Study on the crystal structure of CdInGas, Jpn. J. Appl. Phys. 29 (1990).
[31] N. Cioatera, E.A. Voinea, E. Panaintescu, A. Rolle, S. Somacescu, C.I. Spinu, R. N. Vannier, Changes in structure and electrical conductivity of rare-earth titanate pyrochlores under highly reducing atmosphere, Ceram. Int. 42 (2016) 1492–1500, https://doi.org/10.1016/j.ceramint.2015.09.095.
[32] J.M. Farmer, L.A. Boatner, B.C. Chakoumakos, M.H. Du, M.J. Lance, C.J. Rawn, J. C. Bryan, Structural and crystal chemical properties of rare-earth titanate pyrochlores, J. Alloys Compd. 605 (2014) 63–70, https://doi.org/10.1016/j.jallcom.2014.03.153.
[33] A. Garbout, I. Ben Taazayet-Belgacem, M. Férid, Structural, FT-IR, XRD and Raman scattering of new rare-earth-titanate pyrochlore-type oxides $LnEuTi_2O_7$ (Ln = Gd, Dy), J. Alloys Compd. 573 (2013) 43–52, https://doi.org/10.1016/j.jallcom.2013.03.279.
[34] K. Prashanthi, S. Culubrk, T. Thundat, Applied Surface Science Effect of Annealing Conditions on Structural and Luminescencent Properties of Eu^{3+} -doped $Gd_2Ti_2O_7$ Thin Films Eljka Anti C, vol. 364, 2016, pp. 273–279, https://doi.org/10.1016/j.apsusc.2015.12.137.
[35] C. Ting, Y. Chien, W. Sung, Erbium Doping Effects on the Structural and Infrared Luminescence Properties of $Gd_2Ti_2O_7$ Nanocrystals, vol. 2, 2013, pp. 105–110, https://doi.org/10.1149/2.015306jss.
[36] M.L. Pang, J. Lin, J. Fu, Z.Y. Cheng, Luminescent properties of Gd2Ti2O7:Eu3þ phosphor films prepared by sol – gel process, Mater. Res. Bull. 39 (2004) 1607–1614, https://doi.org/10.1016/j.materresbull.2004.05.016.
[37] Ć. Sanja, V. Lojpur, S.P. Ahrenkiel, J.M. Nedeljkovi, M.D. Drami, Non-contact Thermometry with $Dy^{3þ}$ Doped $Gd_2Ti_2O_7$ Nano-Powders, vol. 170, 2016, pp. 395–400, https://doi.org/10.1016/j.jlumin.2015.06.006.
[38] V. Lojpur, Ć. Sanja, M.D. Drami, Ratiometric luminescence thermometry with different combinations of emissions from $Eu^{3þ}$ doped $Gd_2Ti_2O_7$ nanoparticles. https://doi.org/10.1016/j.jlumin.2015.01.027, 2016, 169, 534-538.
[39] J. Yang, Y. Su, H. Li, X. Liu, Z. Chen, Hydrothermal synthesis and photoluminescence of Ce^{3+} and Tb^{3+} doped $La_2Sn_2O_7$ nanocrystals, J. Alloys Compd. 509 (2011) 8008–8012, https://doi.org/10.1016/j.jallcom.2011.05.082.
[40] P.K. Kulriya, T. Yao, S.M. Scott, S. Nanda, J. Lian, In fl Influence of grain growth on the structural properties of the nanocrystalline $Gd_2Ti_2O_7$, J. Nucl. Mater. 487 (2017) 373–379, https://doi.org/10.1016/j.jnucmat.2017.02.032.
[41] S.K. Gupta, P.S. Ghosh, C. Reghukumar, N. Pathak, R.M. Kadam, Experimental and theoretical approach to account for green luminescence from $Gd_2Zr_2O_7$ pyrochlore: exploring the site occupancy and origin of host-dopant energy transfer in $Gd_2Zr_2O_7:Eu^{3+}$, RSC Adv. 6 (2016) 44908–44920, https://doi.org/10.1039/c6ra05113h.
[42] Y. Yang, Y. Shi, J. Duan, K. Lu, G. Cheng, Y. Zhao, Photoluminescence enhancement of $Gd_2Zr_2O_7$: Eu^{3+} red phosphor sensitized by co-doped Al^{3+} ions, Ceram. Int. (2021) 1–7, https://doi.org/10.1016/j.ceramint.2021.01.171.
[43] G.M. Mustafa, S. Atiq, S.K. Abbas, S. Riaz, S. Naseem, Tunable structural and electrical impedance properties of pyrochlores based Nd doped lanthanum zirconate nanoparticles for capacitive applications, Ceram. Int. 44 (2018) 2170–2177, https://doi.org/10.1016/j.ceramint.2017.10.172.
[44] S.K. Sharma, V. Grover, A.K. Tyagi, D.K. Avasthi, U.B. Singh, P.K. Kulriya, Probing the temperature effects in the radiation stability of $Nd_2Zr_2O_7$ pyrochlore under swift ion irradiation, Materialia 6 (2019), 100317.
[45] N.J. Hess, B.D. Begg, S.D. Conradson, D.E. McCready, P.L. Gassman, W.J. Weber, Spectroscopic investigations of the structural phase transition in $Gd_2(Ti_{1-y}Zr_y)_2O_7$ pyrochlores, J. Phys. Chem. B 106 (2002) 4663–4677, https://doi.org/10.1021/jp014285t.
[46] Z.Y. Ȝ, D. Linghong, P. Xinling, Z. Weifeng, Influence of annealing temperature on luminescent properties of Eu^{3+}/V^{5+} co-doped nanocrystalline Gd_2Ti_2O powders, J. Rare Earths 27 (2009) 900–904, https://doi.org/10.1016/S1002-0721(08)60359-2.
[47] M. Glerup, O.F. Nielsen, F.W. Poulsen, The structural transformation from the pyrochlore structure , $A_2B_2O_7$, to the fluorite structure , AO_2 , studied by Raman spectroscopy and defect chemistry, Modeling 32 (2001) 25–32, https://doi.org/10.1006/jssc.2000.9142.
[48] A. Garbout, T. Turki, M. Férid, Structural and photoluminescence characteristics of Sm^{3+} activated $RE_2Ti_2O_7$ (RE = Gd , La) as orange-red emitting phosphors, J. Lumin. 196 (2018) 326–336, https://doi.org/10.1016/j.jlumin.2017.12.066.
[49] G. Sattonnay, S. Moll, L. Thoḿ, C. Decorse, C. Legros, P. Simon, J. Jagielski, I. Jozwik, I. Monnet, Phase transformations induced by high electronic excitation in ion-irradiated $Gd_2(Zr_xTi_{1-x})_2O_7$ pyrochlores, J. Appl. Phys. 108 (2010), https://doi.org/10.1063/1.3503452.
[50] H. Gupta, J. Singh, R.N. Dutt, S. Ojha, S. Kar, R. Kumar, V.R. Reddy, F. Singh, Defect-induced photoluminescence from gallium-doped zinc oxide thin films:

influence of doping and energetic ion irradiation, Phys. Chem. Chem. Phys. 21 (2019) 15019–15029, https://doi.org/10.1039/c9cp02148e.

[51] J. Jassi, V. Biju, A. Shajan, Structural and optical properties of zinc oxide nanoparticles, AIP Conf. Proc. 2263 (2020) 503–508, https://doi.org/10.1063/5.0017263.

[52] P.A.M. Berdowski, M.J.J. Lammers, G. Blasse, $^5D_3$-$^5D_4$ cross-relaxation of Tb^{3+} in α-GdOF, Chem. Phys. Lett. 113 (1985) 387–390, https://doi.org/10.1016/0009-2614(85)80387-0.

arXiv:2205.12411v1 [cs.LG] 24 May 2022

Linear Connectivity Reveals Generalization Strategies

Jeevesh Juneja
Delhi Technological University
jeeveshjuneja@gmail.com

Rachit Bansal
Delhi Technological University
racbansa@gmail.com

Kyunghyun Cho
New York University
kyunghyun.cho@nyu.edu

João Sedoc
New York University
jsedoc@stern.nyu.edu

Naomi Saphra
New York University
nsaphra@nyu.edu

Abstract

It is widely accepted in the mode connectivity literature that when two neural networks are trained similarly on the same data, they are connected by a path through parameter space over which test set accuracy is maintained. Under some circumstances, including transfer learning from pretrained models, these paths are presumed to be linear. In contrast to existing results, we find that among text classifiers (trained on MNLI, QQP, and CoLA), some pairs of finetuned models have large barriers of increasing loss on the linear paths between them. On each task, we find distinct clusters of models which are linearly connected on the test loss surface, but are disconnected from models outside the cluster—models that occupy separate basins on the surface. By measuring performance on specially-crafted diagnostic datasets, we find that these clusters correspond to different generalization strategies: one cluster behaves like a bag of words model under domain shift, while another cluster uses syntactic heuristics. Our work demonstrates how the geometry of the loss surface can guide models towards different heuristic functions.

1 Introduction

Modern training methods are capable of discovering high-performance parameters for neural networks on a variety of tasks. Although models trained with similar procedures tend to exhibit similar in-domain (ID) performance on these tasks, they exhibit diverse decision boundaries [1]. In particular, models with similar performance can make significantly different judgments when presented with examples that fall far from the training data manifold [43]. In computer vision, these discussions of generalization often focus on visualizations of decision boundaries learnt by the different models.

In NLP, generalization behavior is instead usually characterized structurally through the use of diagnostic sets. Previous studies of model behavior on out-of-distribution (OOD) linguistic structures show that finetuned models can exhibit variation in compositional generalization and performance on diagnostic sets [31, 56]. For example, in natural language inference tasks, some models seem to deploy strategies during OOD generalization that incorporate no position information at all [32]. To the best of our knowledge, these different generalization behaviors have never been linked to the geometry of the loss surface. In order to explore how barriers in the loss surface expose a model's generalization strategy, we consider the case of text classification. We focus in particular on Natural Language Inference [NLI; 50, 3], as well as paraphrase and grammatical acceptability tasks.

We find that NLI models tend to rely on one of two strategies, both of which exhibit similar loss on the test set. We characterize these strategies as, roughly, **syntax-aware** and **syntax-unaware**.

Preprint. Under review.

They fall into two respective basins; the syntax-aware basin contains functions that tend to rely on heuristics around the behavior of constituents, while functions in the syntax-unaware basin rely on lexical bag-of-words heuristics. We find that in NLI, paraphrase, and grammaticality judgment tasks, models that perform similarly on the same diagnostic sets are linearly connected without barriers on the ID loss surface, but they tend to be disconnected from models with different generalization behavior. Our main contributions are:

- We formally define the notion of an ϵ-convex basin to describe sets of weights that can linearly connect to the same low-loss points without crossing large loss surface barriers (Section 3.1).
- We develop a metric based on linear mode connectivity, the **convexity gap** (Section 4), and an accompanying method for clustering models into basins (Section 4.1). In contrast with existing work in computer vision [35], we find that transfer learning can lead to different basins over different finetuning runs (Section 3).
- We align the basins to specific generalization behaviors. In NLI (Section 2.1), they correspond to a preference for either syntactic or lexical overlap heuristics (Section 4). On a paraphrase task (Section 2.2), the basins likewise split on behavior under word order permutation (Section 4.1.1). On a linguistic acceptability task, they reveal the ability to classify linguistic phenomena unseen during finetuning (Appendix A).
- We confirm that these basins trap a portion of finetuning runs, which become increasingly disconnected from the other models as they train (Section 4.2). Based on this behavior, it may be possible to predict final heuristics from early connectivity.

2 Identifying generalization strategies

Finetuning on standard GLUE [45] datasets often leads to models that perform similarly on in-domain (ID) test sets [40]. To evaluate the functional differences between these models, we need to evaluate their generalization to diagnostic datasets. In this paper, we study the variation of performance on these existing diagnostic datasets. We will call models with poor performance on the generalization set **heuristic models** while those with high performance will be **generalizing models**. We study three tasks with diagnostic sets: NLI, paraphrase, and linguistic acceptability (the latter in Appendix A).

2.1 Natural Language Inference

NLI is a common testbed for NLP models. This binary classification task poses a challenge in modeling both syntax and semantics. The input to an NLI model is a pair of sentences such as:

- *Premise:* The dog scared the cat.
- *Hypothesis:* The cat was scared by the dog.

Here, the label is positive or **entailment**, because the hypothesis can be inferred from the premise. If the hypothesis were, "The dog was scared by the cat", the label would be negative or **non-entailment**. We use the MNLI [50] corpus, and inspect loss surfaces on the "matched" validation set.[1]

NLI models often "cheat" by relying on heuristics, such as overlap between either individual lexical items or syntactic constituents shared by the premise and hypothesis. If a model relies on lexical overlap, both the entailed and non-entailed examples above might be given positive labels, because all three sentences contain "scared", "dog", and "cat". McCoy et al. [32] responded to these shortcuts by creating HANS, a diagnostic set of sentence pairs that violate such heuristics:

- **Lexical overlap (HANS-LO):** Entails any hypothesis containing the same words as the premise.
- **Subsequence:** Entails any hypothesis containing contiguous sequences of words from the premise.
- **Constituent:** Entails any hypothesis containing syntactic subtrees from the premise.

Unless otherwise specified, we use the non-entailing HANS subsets for measuring reliance on heuristics, so higher accuracy on HANS-LO indicates less reliance on lexical overlap.

[1] An unmatched validation set is also available, which includes different sources and topics from the training set. The test set labels are not public for MNLI or QQP, so we use the validation set only.

2.2 Paraphrase

Quora Question Pairs [QQP; 47] is a common paraphrase corpus. The dataset contains 400k pairs of questions that have been annotated with a binary value indicating whether they are duplicates. We use the PAWS-QQP [54] dataset as the diagnostic set for identifying different generalization behaviors learnt by the models. PAWS-QQP contains QQP sentences that have been permuted in order to construct pairs that can mean different things, even though they have the same words. In other words, they are pairs that may violate a lexical overlap heuristic.

2.3 Experimental details

All models are initialized from `bert-base-uncased`,[2] with a linear classification head. Each model has a random seed determining both the initialisation of the linear layer and the data order. QQP models are trained with Google's recommended default hyperparameters (details in Appendix F). MNLI models are those provided by McCoy et al. [31]. Both have been trained using Google's original trainer.[3]

2.3.1 Computational resources

Finetuning 100 QQP models cost around 500 GPU-hours and 48 CoLA models cost around 10 GPU-hours. Each 100×100 interpolation and evaluation consumed 114 GPU hours; these were performed for both QQP (at three stages during finetuning, for Fig. 7) and MNLI. The 48×48 interpolation for CoLA cost about 28 GPU-hours. The experiments add up to a total cost of approximately 994 GPU-hours on a mix of NVIDIA RTX8000 and V100 nodes.

3 Linear Mode Connectivity

Models discovered by SGD are generally connected by paths over which the loss is maintained [6]. If we limit such paths to be linear, however, connectivity is no longer guaranteed. We may still find, however, that two parameter settings θ_A and θ_B, which achieve equal loss, can be connected by linear interpolation without any increase in loss [12]. In other words, loss $\mathcal{L}(\theta_\alpha; X_{\text{train}}, Y_{\text{train}}) \leq \mathcal{L}(\theta_A; X_{\text{train}}, Y_{\text{train}})$ and $\mathcal{L}(\theta_\alpha; X_{\text{train}}, Y_{\text{train}}) \leq \mathcal{L}(\theta_B; X_{\text{train}}, Y_{\text{train}})$ in each parameter setting θ_α defined by any scalar $0 \leq \alpha \leq 1$:

$$\theta_\alpha = \alpha\theta_A + (1-\alpha)\theta_B \tag{1}$$

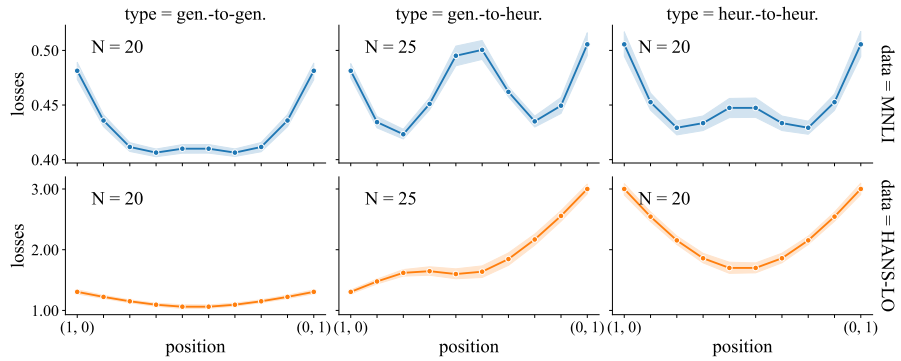
Frankle et al. [9] tested linear mode connectivity between models initialized from a single pruned network. They found that the resulting models achieved the same performance as the entire un-pruned network only when they were linearly connected. This result suggests that generalization, or at least high performance, is closely tied to the linear mode connectivity of the models in question. This implied result is further supported by Entezari et al. [7], which found a larger barrier between models when they exhibited higher test error. Neyshabur et al. [35] even described linear mode connectivity as a crucial component of transfer learning, finding that finetuned models initialized from the same pretrained model will be in the same linearly connected basin, in contrast to models trained from scratch, which exhibit barriers even when initialized from the same random weights.

Our results complicate the narrative around linear mode connectivity (Fig. 1(a)). While we find that models with high performance on OOD data were indeed linearly connected to each other, models with low performance were also linearly connected to each other. It seems that the heuristic and generalizing models occupy two different linear basins, with barriers in the ID loss surface between models in each of these two basins.

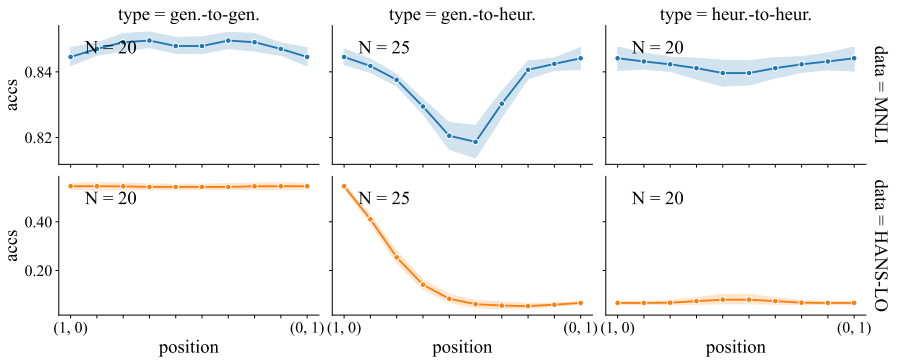
HANS performance during interpolation: It is clear from Fig. 1(a) that, when interpolating between LO-heuristic models, HANS-LO loss significantly improves further from the end points. This finding implies that the heuristic basin does contain more syntax-aware models. In contrast,

[2] `https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip`

[3] `https://github.com/google-research/bert`



(a) MNLI and HANS-LO loss variation during linear interpolation.



(b) MNLI and HANS-LO accuracy variation during linear interpolation.

Figure 1: Performance during linear interpolation between pairs of models taken from the 5 top (gen.) and 5 bottom (heur.) in HANS-LO accuracy. Heuristic models tend to be poorly connected to the rest of the surface, although they are well connected to each other. N indicates number of pairs. Position on the x-axis indicates the value of α during interpolation.

the syntax-aware basin shows only a slight improvement in heuristic loss during interpolation, even though the improvement in ID test loss is more substantial than in the position-unaware basin. However, we can see from Fig. 1(b) that low losses on interpolated models do not always translate to consistently higher accuracy, although interpolated models that fall on barriers of elevated test loss do lead to lower ID accuracy.

Connections over 2 dimensions: To understand the loss topography better, we present planar views of these models using code[4] from Benton et al. [1]. We see how, in the plane covering the heuristic and generalizing models, a central barrier intrudes on the linear connecting edge between heuristic and generalizing models (Fig. 2). On the other hand, the heuristic and generalizing models each occupy separate planes that exhibit only a small central barrier. Visibly, this barrier is smallest for the perimeter composed of generalizing models and largest for the mixed perimeter. These topographies motivate the following notion of an ϵ-convex basin.

3.1 Convex basins

Inspired by Entezari et al. [7], we define our notion of a basin in order to formalize types of behavior during linear interpolation. In contrast to their work, however, our goal is not to identify a largely stable region (the bottom of a basin). Instead, we are interested in whether a set of models are connected to the same low-loss points by a linear path of non-increasing loss. This motivation yields the following definition.

[4] `https://github.com/g-benton/loss-surface-simplexes`

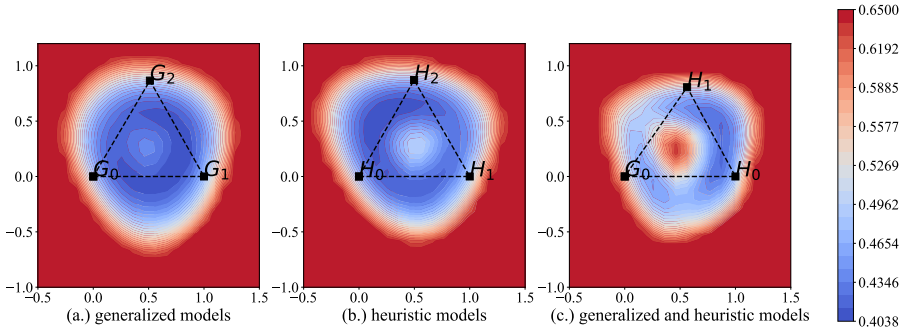


Figure 2: Planar views of simplices connecting NLI models in the MNLI matched validation set loss surface. The points $G_{0...2}$ and $H_{0...2}$ denote the generalizing and heuristic models respectively.

For resolution $\epsilon \geq 0$, we define an **ϵ-convex basin** as a convex set S, such that, for any set of points $w_1, w_2, ..., w_k \in S$ and any set of coefficients $\alpha_1 \dots \alpha_n \geq 0$ where $\sum_k \alpha_k = 1$, a relaxed form of Jensen's inequality holds:

$$\mathcal{L}(\sum_{k=1}^{n} \alpha_k w_k) \leq \epsilon + \sum_{k=1}^{n} \alpha_k \mathcal{L}(w_k) \tag{2}$$

In particular, we say that a set of trained models $\theta_1, .., \theta_n$ form an ϵ-convex basin if the region inside their convex hull is an ϵ-convex basin. In the case where $\epsilon = 0$, we can equivalently claim that the behavior of $\mathcal{L}$ in the region is convex.

This definition satisfies our stated motivation because any two points within an ϵ-convex basin are connected linearly with monotonically (within ϵ) decaying loss to the same minima within that basin. While such a linear path does not strictly describe a typical SGD optimization trajectory, such linear interpolations are frequently used to analyze optimization behaviors like module criticality [35, 2]. One justification of this practice is that much of the oscillation and complexity of the training trajectory is constrained to short directions of the solution manifold [19, 53, 30], while large scale views of the landscape describe smoother and more linear trajectories [12, 8]. In this paper, we use this definition of ϵ-convexity to describe linearly connected points as existing in the same basin. However, pairwise non-increasing linear connections are a necessary but not sufficient condition for all convex combinations to be non-increasing, with counterexamples in Fig. 2.

4 The Convexity Gap

One possibility we argue against is that the increasing loss between models with different heuristics is actually an effect of sharper minima in the heuristic case. There is a significant body of work on the controversial [5] association between wider minima and generalization in models [26, 21, 17]. Prior work shows that minima forced to memorize a training set without generalizing on a test set exist in very sharp basins [18], but these basins are discovered only by directly pessimizing on the test set. It is less clear whether width is predictive of generalization across models trained conventionally.

On the contrary, we find that clusters of linearly connected models are far better predictors of generalization behavior than optimum width is. To identify such clusters, we define a metric based on linear mode connectivity, the **convexity gap** ($\mathfrak{CG}$), and use it to perform spectral clustering.

Entezari et al. [7] define a barrier's height on a linear path from θ_1 to θ_2 as:

$$\mathfrak{BH}(\theta_1, \theta_2) = \sup_{\alpha} [\mathcal{L}(\alpha\theta_1 + (1-\alpha)\theta_2) - (\alpha\mathcal{L}(\theta_1) + (1-\alpha)\mathcal{L}(\theta_2))] \qquad \alpha \in [0,1] \tag{3}$$

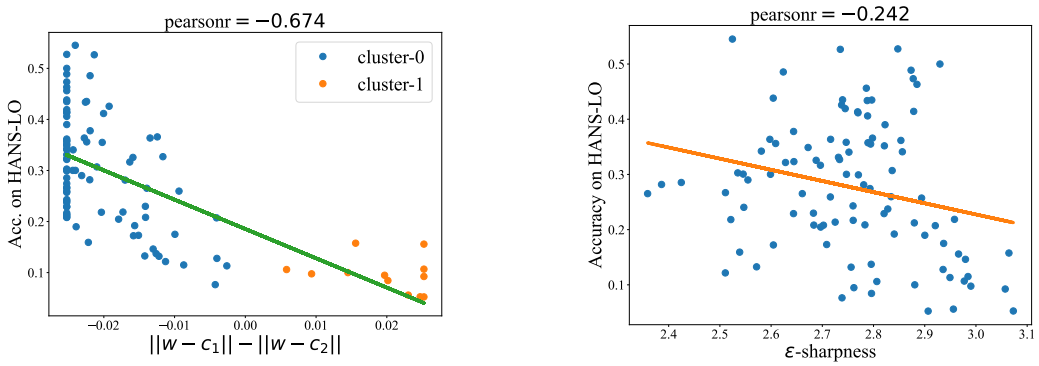


Figure 3: Using features of the ID loss landscape to predict reliance on lexical overlap heuristics. Least squares fit shown. (a) HANS-LO performance vs difference between distances from the centroids c_1, c_2 of the $\mathfrak{CG}$-based clusters. (b) HANS-LO performance vs ϵ-sharpness.

We define the convexity gap on a linear path from θ_1 to θ_2 as the maximum possible barrier height on any sub-segment of the linear path joining θ_1 and θ_2. Mathematically,

$$\mathfrak{CG}(\theta_1, \theta_2) = \sup_{\gamma,\beta} \mathfrak{BH}(\gamma\theta_1 + (1-\gamma)\theta_2, \beta\theta_1 + (1-\beta)\theta_2) \qquad \gamma, \beta \in [0, 1] \tag{4}$$

Under this definition, an ϵ-convex basin has a convexity gap of at most ϵ within the basin (proof in Appendix B). Our metric exposes stronger clustering patterns than either area under the curve (AUC) or $\mathfrak{BH}$ (evidence in Appendix I).

4.1 Clustering

The basins that form from this distance metric are visible based on connected sets of models in the distance heatmap (Fig. 4(a)). To quantify basin membership into a prediction of HANS performance, we perform spectral clustering, with the distances between points defined as $\mathfrak{CG}$-distance on the ID loss surface. Using the difference between distances from each cluster centroid $\|w - c_1\|$ and $\|w - c_2\|$, we see a significantly larger correlation with HANS performance (Fig. 3(a)), compared to a baseline of the model's ϵ-sharpness (Fig. 3(b)). Furthermore, a linear regression model offers significantly better performance based on spectral clustering membership than sharpness.[5]

In Fig. 4(b), we see the heuristic that defines the larger cluster: constituent overlap. Models that perform well on constituent overlap diagnostic sets tend to fall in the basin containing models biased towards lexical overlap. This behavior characterizes the two basins: the larger basin is syntax-aware (tending to acquire heuristics that require awareness of constituent structure), while the smaller basin is syntax-unaware (acquiring heuristics that rely only on unordered sets of words).[6]

Distributions of the clusters: We find (Fig. 5) that $\mathfrak{CG}$-based cluster membership accounts for some of the heavy tail of performance on HANS-LO, supporting the claim that the convex basins on the loss surface differentiate generalization strategies. However, the distribution of ID performance also differs significantly between clusters, so we may choose to relate basin membership to ID behavior instead of OOD. We discuss this alternative framing further in Appendix H.

[5]The connection between cluster membership and generalization performance is continuous. That is, those members of the syntax-aware cluster which are closer to the cluster boundary behave a little more like bag-of-words models, rather than behaving like typical syntax-aware models. The existence of marginally syntax-aware models is unsurprising; we already know that there are models in the heuristic basin that are slightly more syntax-aware, because they are visible at midpoints during interpolation between heuristic models (Fig. 1(a)).

[6]It is not enough to declare one basin to be merely *position* aware, as reliance on subsequence overlap is a poor predictor of basin membership (Appendix C).

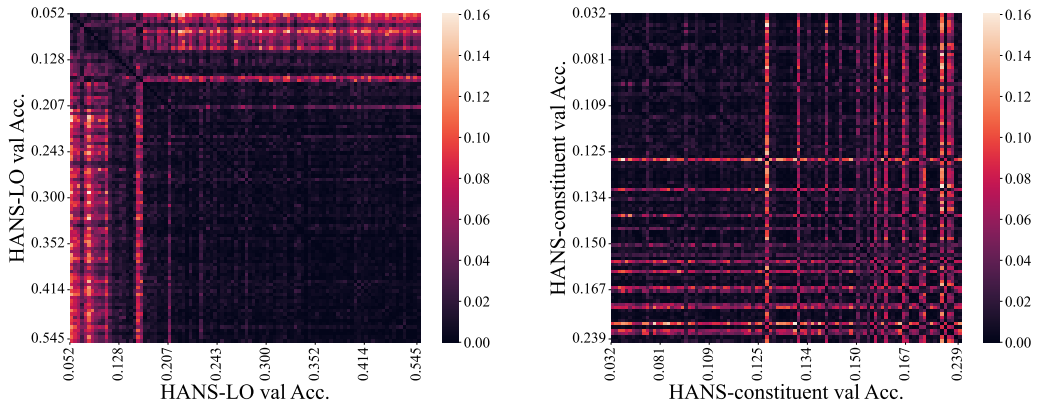


Figure 4: Color indicates $\mathfrak{CG}$ distance between the x- and y-axis NLI models. (a) Models sorted by increasing performance on HANS-LO. (b) Sorted by increasing performance on HANS-constituent.

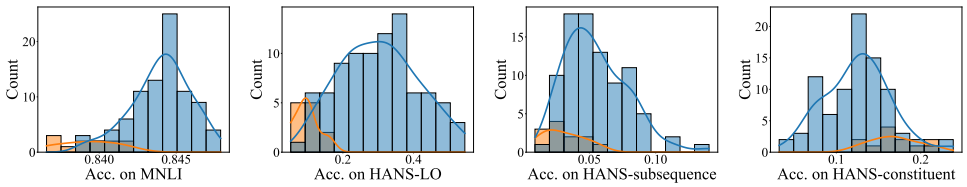


Figure 5: Histogram of accuracy scores of the different NLI clusters.

4.1.1 QQP

On QQP, we similarly find strong differentiation between distinct clusters of linearly connected models (Fig. 6(a)). As in our procedure on NLI, we study the distance from the centroids formed by spectral clustering with $\mathfrak{CG}$ defining the distance metric, and find cluster membership to be a strong predictor of generalization on PAWS-QQP (Fig. 6(b)).

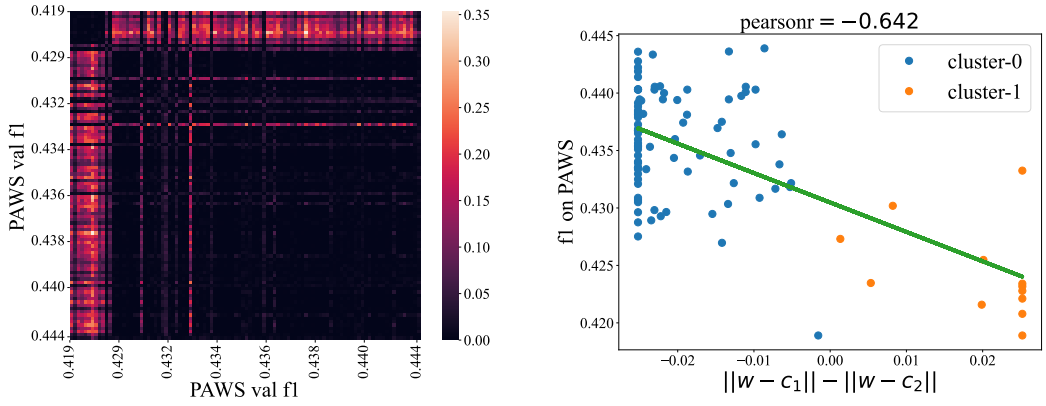


Figure 6: (a) $\mathfrak{CG}$ heatmap with QQP models sorted according to f1 score on PAWS-QQP. (b) QQP cluster membership is highly predictive of PAWS-QQP performance. Least squares fit shown.

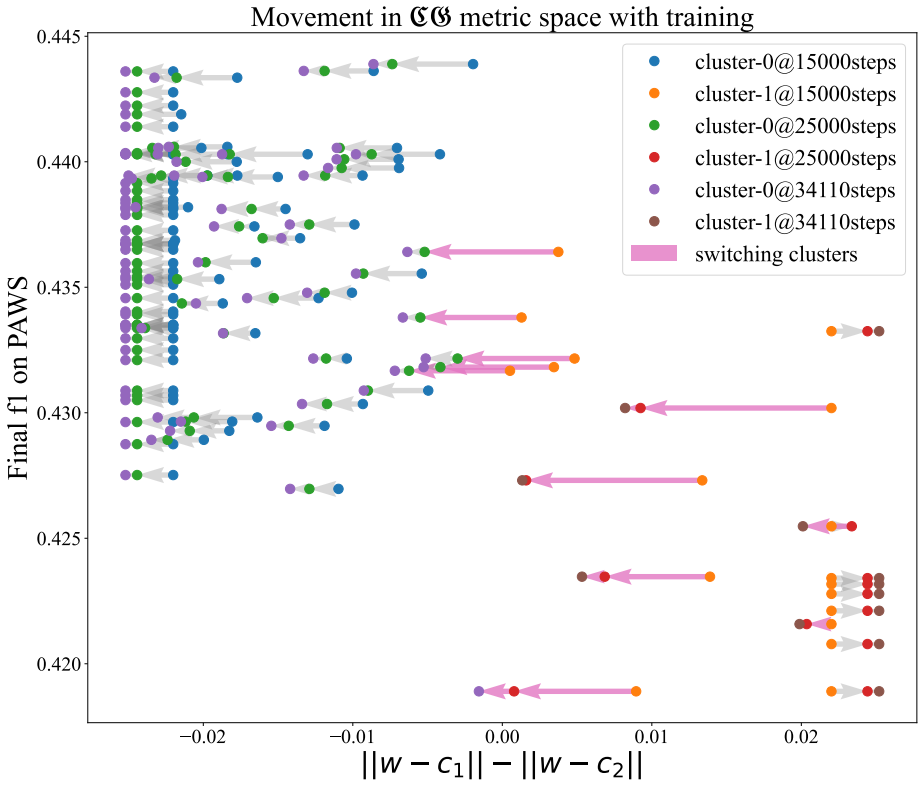


Figure 7: Movement between clusters during finetuning. Syntax-unaware models near the cluster boundary move towards the syntax-aware cluster, but central models in both clusters move closer to their respective centroids.

4.2 Generalization basins trap training trajectories

At this point we have aligned different basins with particular generalization behaviors, but a skeptic may propose that the smaller cluster contains models which have simply not trained long enough. Under this conjecture, continuing to train them would eventually place them in the larger, syntax-aware cluster. We find that this is not the case, as shown in Fig. 7. Marginal members of the smaller cluster, those which fall closer to the cluster boundary, may drift towards the larger cluster later in training. However, models that are more central to the cluster actually become increasingly solidified in their cluster membership later in training.

These results have two important implications. First, they confirm that these basins constitute distinct local minima which can trap optimization trajectories. Additionally, our findings suggest that early on in training, we can detect whether a final model will follow a particular heuristic. The latter conclusion is crucial for any future work towards practical methods of directing or detecting desirable generalization strategies during training [20].

5 Related Work

Connectivity of the Loss Landscape: Draxler et al. [6] demonstrated that pairs of trained models are generally connected to each other by nonlinear paths of constant train and test loss. Garipov et al. [10] and Fort and Jastrzebski [8] conceptualized these paths as high dimensional volumes connecting entire sets of solutions. Goodfellow et al. [12] explored linear connections between model pairs by interpolating between them, while [7] suggested that in general, models that seem linearly disconnected may be connected after permuting their neurons. Our work explores models that are linearly connected without permutation, and is most closely linked with Neyshabur et al. [35], which found that models initialized from a shared pretrained checkpoint are linearly connected. Frankle

et al. [9] also found that pruned initial weights can lead to trained models that interpolate with low loss, if the pruned model generalizes well. These results, unlike ours, focus on image classification.

Generalization: Diagnostic challenge sets are a common method for evaluating generalization behavior in a number of modalities. Imagenet has a number of corresponding diagnostic sets, from naturally-hard examples [25, 15] to perturbed images [14] to sketched renderings of objects [46]. In NLP, diagnostic sets frequently focus on compositionality and grammatical behavior such as inflection agreement [24, 41]. Diagnostic sets in NLP can be based on natural adversarial data [13, 27], constructed by perturbations and rephrases [39, 11], or generated artificially from templates [38, 55]. NLI models in particular often fail on diagnostic sets with permuted word order [22, 34, 32], often exposing a lack of syntactic behavior. Our case study is particularly informative because interpretability of NLI datasets reveals classes of generalization strategies.

Function Diversity: Initializing from a single pretrained model can produce a range of in-domain generalization behaviors [4, 36, 40, 33]. However, variation on performance in diagnostic sets is even more substantial, from social biases [40] to unusual paraphrases [31, 56]. Benton et al. [1] found that a wide variety of decision boundaries were expressed within a low-loss volume, and Somepalli et al. [43] further found that there is diversity in boundaries during OOD generalization, far off the data manifold. Our work contributes to diversity in generalization by linking sets of models that share low dimensional subspaces to particular OOD generalization behavior. We also expand beyond visualizing decision boundaries by focusing on the geography of basins aligned with specific interpretable strategies, particularly syntax-aware heuristics.

6 Discussion and Future Work

Future research may focus on evaluating the strength and nature of the prior distribution over basins. Transfer learning priors range from the high entropy distribution provided by training from scratch to those pretrained models which seem to consistently select a single basin [35]. Possible influences on basin selection, and therefore on generalization strategies, may include length of pretraining [49], data scheduling, and architecture selection [43]. The strength of a prior towards particular basins may be not only linked to training procedure, but also strongly related to the availability of features in the pretrained representations [29, 16, 42]. This aspect of transfer learning is a particularly ripe area for theoretical work in optimization.

Another possible direction for future work is to identify early on whether a basin has a desirable generalization strategy. For example, can we detect early on if a Transformer model is unable to use position information, instead acting as a bag of words? It is particularly promising that $\mathfrak{CG}$ in training loss is highly predictive of OOD generalization even without any validation set (Appendix G), so the only investigation needed is of the loss surface during training. Furthermore, a model tends to stay in the same cluster in early and late training (Section 4.2), so final heuristics should be easy to predict from earlier basin membership.

The split between generalization strategies can potentially explain results from the bimodality of CoLA models [33] to wide variance on NLI diagnostic sets [31]. Because weight averaging can find parameter settings that fall on a barrier, we may even explain why weight averaging, which tends to perform well on vision tasks, fails in text classifiers [52]. Future work that distinguishes generalization strategy basins could improve the performance of such weight ensembling methods.

Because these methods rely on a high likelihood of linear mode connectivity, they are particular to transfer learning [35] or other settings with a similar connectivity bias [9]. We could also study pretrained models themselves or training from scratch, but only if we first align the neurons [7, 44] so that the vectors being averaged match in their role within the function space.

Within NLP, many tasks and accompanying generalization strategies can be further explored. There are general tests of inductive bias towards grammatical behavior [49], as well as other diagnostic sets. Outside of text classification, sequence-to-sequence tasks provide a number of possible tests for generalization behavior. For language modeling itself, we can consider biases towards subject-verb number agreement or other ways of matching inflection [27, 13, 37]. OOD generalization tests are also available for seq2seq tasks like semantic parsing [23].

While we focus on text classification due to the availability of diagnostic datasets that test structural generalization, other modalities such as vision may have distinct generalization strategies. In some domain shift scenarios [25, 46, 15, 14], OOD generalization behavior may therefore form basins. In current work, however, barriers are not observed in vision during transfer [35, 52]. Is that because *all* vision models achieve a basin with good structural generalization—or none do?

7 Conclusions

In one view of transfer learning, a pretrained model may select a prior over generalization strategies by favoring a particular basin. Neyshabur et al. [35] found that models initialized from the same pretrained weights are linearly connected, suggesting that basin selection is a key component of transfer learning. We find that a pretrained model may not commit exclusively to a single basin, but instead favor a small set of them. Furthermore, we find that linear connectivity can indicate shared generalization strategies under domain shift, as evidenced by results on NLI, paraphrase, and linguistic acceptability tasks.

References

[1] G. W. Benton, W. J. Maddox, S. Lotfi, and A. G. Wilson. Loss Surface Simplexes for Mode Connecting Volumes and Fast Ensembling. *arXiv:2102.13042 [cs, stat]*, Feb. 2021. URL http://arxiv.org/abs/2102.13042. arXiv: 2102.13042.

[2] N. Chatterji, B. Neyshabur, and H. Sedghi. The intriguing role of module criticality in the generalization of deep networks. In *International Conference on Learning Representations*, 2020. URL https://openreview.net/forum?id=S1e4jkSKvB.

[3] T. F. Consortium, R. Cooper, D. Crouch, J. V. Eijck, C. Fox, J. V. Genabith, J. Jaspars, H. Kamp, D. Milward, M. Pinkal, M. Poesio, S. Pulman, T. Briscoe, H. Maier, and K. Konrad. Using the framework, 1996.

[4] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *ArXiv*, abs/1810.04805, 2019.

[5] L. Dinh, R. Pascanu, S. Bengio, and Y. Bengio. Sharp Minima Can Generalize For Deep Nets. *arXiv:1703.04933 [cs]*, May 2017. URL http://arxiv.org/abs/1703.04933. arXiv: 1703.04933.

[6] F. Draxler, K. Veschgini, M. Salmhofer, and F. A. Hamprecht. Essentially No Barriers in Neural Network Energy Landscape. *arXiv:1803.00885 [cs, stat]*, Feb. 2019. URL http://arxiv.org/abs/1803.00885. arXiv: 1803.00885.

[7] R. Entezari, H. Sedghi, O. Saukh, and B. Neyshabur. The Role of Permutation Invariance in Linear Mode Connectivity of Neural Networks. *arXiv:2110.06296 [cs]*, Oct. 2021. URL http://arxiv.org/abs/2110.06296. arXiv: 2110.06296.

[8] S. Fort and S. Jastrzebski. Large Scale Structure of Neural Network Loss Landscapes. In *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. URL https://proceedings.neurips.cc/paper/2019/hash/48042b1dae4950fef2bd2aafa0b971a1-Abstract.html.

[9] J. Frankle, G. K. Dziugaite, D. Roy, and M. Carbin. Linear Mode Connectivity and the Lottery Ticket Hypothesis. In *Proceedings of the 37th International Conference on Machine Learning*, pages 3259–3269. PMLR, Nov. 2020. URL https://proceedings.mlr.press/v119/frankle20a.html. ISSN: 2640-3498.

[10] T. Garipov, P. Izmailov, D. Podoprikhin, D. Vetrov, and A. G. Wilson. Loss Surfaces, Mode Connectivity, and Fast Ensembling of DNNs. *arXiv:1802.10026 [cs, stat]*, Oct. 2018. URL http://arxiv.org/abs/1802.10026. arXiv: 1802.10026.

[11] M. Glockner, V. Shwartz, and Y. Goldberg. Breaking NLI systems with sentences that require simple lexical inferences. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 650–655, Melbourne, Australia, July 2018. Association for Computational Linguistics. doi: 10.18653/v1/P18-2103. URL https://aclanthology.org/P18-2103.

[12] I. J. Goodfellow, O. Vinyals, and A. M. Saxe. Qualitatively characterizing neural network optimization problems. *arXiv:1412.6544 [cs, stat]*, May 2015. URL http://arxiv.org/abs/1412.6544. arXiv: 1412.6544.

[13] K. Gulordava, P. Bojanowski, E. Grave, T. Linzen, and M. Baroni. Colorless green recurrent networks dream hierarchically. *arXiv:1803.11138 [cs]*, Mar. 2018. URL http://arxiv.org/abs/1803.11138. arXiv: 1803.11138.

[14] D. Hendrycks and T. Dietterich. Benchmarking neural network robustness to common corruptions and perturbations. *Proceedings of the International Conference on Learning Representations*, 2019.

[15] D. Hendrycks, K. Zhao, S. Basart, J. Steinhardt, and D. Song. Natural adversarial examples. *CVPR*, 2021.

[16] K. L. Hermann and A. K. Lampinen. What shapes feature representations? Exploring datasets, architectures, and training. *arXiv:2006.12433 [cs, stat]*, Oct. 2020. URL http://arxiv.org/abs/2006.12433. arXiv: 2006.12433.

[17] S. Hochreiter and J. Schmidhuber. Flat minima. *Neural Computation*, 9:1–42, 1997.

[18] W. R. Huang, Z. Emam, M. Goldblum, L. Fowl, J. K. Terry, F. Huang, and T. Goldstein. Understanding Generalization through Visualizations. *arXiv:1906.03291 [cs, stat]*, Nov. 2020. URL http://arxiv.org/abs/1906.03291. arXiv: 1906.03291.

[19] S. Jastrzebski, M. Szymczak, S. Fort, D. Arpit, J. Tabor, K. Cho, and K. Geras. The Break-Even Point on Optimization Trajectories of Deep Neural Networks. *arXiv:2002.09572 [cs, stat]*, Feb. 2020. URL http://arxiv.org/abs/2002.09572. arXiv: 2002.09572.

[20] S. Jastrzebski, D. Arpit, O. Astrand, G. B. Kerg, H. Wang, C. Xiong, R. Socher, K. Cho, and K. J. Geras. Catastrophic Fisher Explosion: Early Phase Fisher Matrix Impacts Generalization. In *Proceedings of the 38th International Conference on Machine Learning*, pages 4772–4784. PMLR, July 2021. URL https://proceedings.mlr.press/v139/jastrzebski21a.html. ISSN: 2640-3498.

[21] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang. On large-batch training for deep learning: Generalization gap and sharp minima. *ArXiv*, abs/1609.04836, 2017.

[22] J. Kim, C. Malon, and A. Kadav. Teaching syntax by adversarial distraction. In *Proceedings of the First Workshop on Fact Extraction and VERification (FEVER)*, pages 79–84, Brussels, Belgium, Nov. 2018. Association for Computational Linguistics. doi: 10.18653/v1/W18-5512. URL https://aclanthology.org/W18-5512.

[23] N. Kim and T. Linzen. COGS: A Compositional Generalization Challenge Based on Semantic Interpretation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 9087–9105, Online, Nov. 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-main.731. URL https://aclanthology.org/2020.emnlp-main.731.

[24] N. Kim, R. Patel, A. Poliak, P. Xia, A. Wang, T. McCoy, I. Tenney, A. Ross, T. Linzen, B. Van Durme, S. R. Bowman, and E. Pavlick. Probing what different NLP tasks teach machines about function word comprehension. In *Proceedings of the Eighth Joint Conference on Lexical and Computational Semantics (*SEM 2019)*, pages 235–249, Minneapolis, Minnesota, June 2019. Association for Computational Linguistics. doi: 10.18653/v1/S19-1026. URL https://aclanthology.org/S19-1026.

[25] P. W. Koh, S. Sagawa, H. Marklund, S. M. Xie, M. Zhang, A. Balsubramani, W. Hu, M. Yasunaga, R. L. Phillips, I. Gao, T. Lee, E. David, I. Stavness, W. Guo, B. A. Earnshaw, I. S. Haque, S. Beery, J. Leskovec, A. Kundaje, E. Pierson, S. Levine, C. Finn, and P. Liang. WILDS: A Benchmark of in-the-Wild Distribution Shifts. *arXiv:2012.07421 [cs]*, July 2021. URL http://arxiv.org/abs/2012.07421. arXiv: 2012.07421.

[26] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein. Visualizing the Loss Landscape of Neural Nets. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems 31*, pages 6389–6399. Curran Associates, Inc., 2018. URL http://papers.nips.cc/paper/7875-visualizing-the-loss-landscape-of-neural-nets.pdf.

[27] T. Linzen, E. Dupoux, and Y. Goldberg. Assessing the Ability of LSTMs to Learn Syntax-Sensitive Dependencies. *arXiv:1611.01368 [cs]*, Nov. 2016. URL http://arxiv.org/abs/1611.01368. arXiv: 1611.01368.

[28] I. Loshchilov and F. Hutter. Decoupled weight decay regularization, 2017. URL https://arxiv.org/abs/1711.05101.

[29] C. Lovering, R. Jha, T. Linzen, and E. Pavlick. Predicting inductive biases of pre-trained models. In *ICLR*, 2021.

[30] C. Ma, L. Wu, and L. Ying. The Multiscale Structure of Neural Network Loss Functions: The Effect on Optimization and Origin. *arXiv:2204.11326 [cs]*, Apr. 2022. URL http://arxiv.org/abs/2204.11326. arXiv: 2204.11326.

[31] R. T. McCoy, J. Min, and T. Linzen. Berts of a feather do not generalize together: Large variability in generalization across models with similar test set performance. *ArXiv*, abs/1911.02969, 2020.

[32] T. McCoy, E. Pavlick, and T. Linzen. Right for the Wrong Reasons: Diagnosing Syntactic Heuristics in Natural Language Inference. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 3428–3448, Florence, Italy, July 2019. Association for Computational Linguistics. doi: 10.18653/v1/P19-1334. URL https://aclanthology.org/P19-1334.

[33] M. Mosbach, M. Andriushchenko, and D. Klakow. On the Stability of Fine-tuning BERT: Misconceptions, Explanations, and Strong Baselines. *arXiv:2006.04884 [cs, stat]*, Mar. 2021. URL http://arxiv.org/abs/2006.04884. arXiv: 2006.04884.

[34] A. Naik, A. Ravichander, N. Sadeh, C. Rose, and G. Neubig. Stress test evaluation for natural language inference. In *Proceedings of the 27th International Conference on Computational Linguistics*, pages 2340–2353, Santa Fe, New Mexico, USA, Aug. 2018. Association for Computational Linguistics. URL https://aclanthology.org/C18-1198.

[35] B. Neyshabur, H. Sedghi, and C. Zhang. What is being transferred in transfer learning? In *Advances in Neural Information Processing Systems*, volume 33, pages 512–523. Curran Associates, Inc., 2020. URL https://papers.nips.cc/paper/2020/hash/0607f4c705595b911a4f3e7a127b44e0-Abstract.html.

[36] J. Phang, J. Park, and K. J. Geras. Investigating and Simplifying Masking-based Saliency Methods for Model Interpretability. *arXiv:2010.09750 [cs]*, Oct. 2020. URL http://arxiv.org/abs/2010.09750. arXiv: 2010.09750.

[37] S. Ravfogel, Y. Goldberg, and T. Linzen. Studying the Inductive Biases of RNNs with Synthetic Variations of Natural Languages. *arXiv:1903.06400 [cs]*, Mar. 2019. URL http://arxiv.org/abs/1903.06400. arXiv: 1903.06400.

[38] R. Rudinger, J. Naradowsky, B. Leonard, and B. Van Durme. Gender bias in coreference resolution. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 8–14, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-2002. URL https://aclanthology.org/N18-2002.

[39] I. Sanchez, J. Mitchell, and S. Riedel. Behavior analysis of NLI models: Uncovering the influence of three factors on robustness. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1975–1985, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-1179. URL https://aclanthology.org/N18-1179.

[40] T. Sellam, S. Yadlowsky, I. Tenney, J. Wei, N. Saphra, A. D'Amour, T. Linzen, J. Bastings, I. R. Turc, J. Eisenstein, D. Das, and E. Pavlick. The MultiBERTs: BERT Reproductions for Robustness Analysis. In *ICLR*, Sept. 2021. URL https://openreview.net/forum?id=K0E_F0gFDgA.

[41] R. Sennrich. How grammatical is character-level neural machine translation? assessing MT quality with contrastive translation pairs. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 2, Short Papers*, pages 376–382, Valencia, Spain, Apr. 2017. Association for Computational Linguistics. URL https://aclanthology.org/E17-2060.

[42] H. Shah, K. Tamuly, A. Raghunathan, P. Jain, and P. Netrapalli. The Pitfalls of Simplicity Bias in Neural Networks. *arXiv:2006.07710 [cs, stat]*, Oct. 2020. URL http://arxiv.org/abs/2006.07710. arXiv: 2006.07710.

[43] G. Somepalli, L. Fowl, A. Bansal, P. Yeh-Chiang, Y. Dar, R. Baraniuk, M. Goldblum, and T. Goldstein. Can Neural Nets Learn the Same Model Twice? Investigating Reproducibility and Double Descent from the Decision Boundary Perspective. *arXiv:2203.08124 [cs]*, Mar. 2022. URL http://arxiv.org/abs/2203.08124. arXiv: 2203.08124.

[44] N. J. Tatro, P.-Y. Chen, P. Das, I. Melnyk, P. Sattigeri, and R. Lai. Optimizing Mode Connectivity via Neuron Alignment. *arXiv:2009.02439 [cs, math, stat]*, Nov. 2020. URL http://arxiv.org/abs/2009.02439. arXiv: 2009.02439.

[45] A. Wang, A. Singh, J. Michael, F. Hill, O. Levy, and S. R. Bowman. Glue: A multi-task benchmark and analysis platform for natural language understanding. *ArXiv*, abs/1804.07461, 2018.

[46] H. Wang, S. Ge, Z. Lipton, and E. P. Xing. Learning robust global representations by penalizing local predictive power. In *Advances in Neural Information Processing Systems*, pages 10506–10518, 2019.

[47] Z. Wang, W. Hamza, and R. Florian. Bilateral multi-perspective matching for natural language sentences. *ArXiv*, abs/1702.03814, 2017.

[48] A. Warstadt, A. Singh, and S. R. Bowman. Neural network acceptability judgments. *arXiv preprint arXiv:1805.12471*, 2018.

[49] A. Warstadt, Y. Zhang, X. Li, H. Liu, and S. R. Bowman. Learning Which Features Matter: RoBERTa Acquires a Preference for Linguistic Generalizations (Eventually). In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 217–235, Online, Nov. 2020. Association for Computational Linguistics. URL https://www.aclweb.org/anthology/2020.emnlp-main.16.

[50] A. Williams, N. Nangia, and S. Bowman. A broad-coverage challenge corpus for sentence understanding through inference. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 1112–1122. Association for Computational Linguistics, 2018. URL http://aclweb.org/anthology/N18-1101.

[51] T. Wolf, L. Debut, V. Sanh, J. Chaumond, C. Delangue, A. Moi, P. Cistac, T. Rault, R. Louf, M. Funtowicz, J. Davison, S. Shleifer, P. von Platen, C. Ma, Y. Jernite, J. Plu, C. Xu, T. Le Scao, S. Gugger, M. Drame, Q. Lhoest, and A. Rush. Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 38–45, Online, Oct. 2020. Association for Computational Linguistics. doi: 10.18653/v1/2020.emnlp-demos.6. URL https://aclanthology.org/2020.emnlp-demos.6.

[52] M. Wortsman, G. Ilharco, S. Y. Gadre, R. Roelofs, R. Gontijo-Lopes, A. S. Morcos, H. Namkoong, A. Farhadi, Y. Carmon, S. Kornblith, and L. Schmidt. Model soups: averaging weights of multiple fine-tuned models improves accuracy without increasing inference time. *arXiv:2203.05482 [cs]*, Mar. 2022. URL http://arxiv.org/abs/2203.05482. arXiv: 2203.05482.

[53] C. Xing, D. Arpit, C. Tsirigotis, and Y. Bengio. A Walk with SGD. *arXiv:1802.08770 [cs, stat]*, May 2018. URL http://arxiv.org/abs/1802.08770. arXiv: 1802.08770.

[54] Y. Zhang, J. Baldridge, and L. He. PAWS: Paraphrase Adversaries from Word Scrambling. In *Proc. of NAACL*, 2019.

[55] J. Zhao, T. Wang, M. Yatskar, V. Ordonez, and K.-W. Chang. Gender bias in coreference resolution: Evaluation and debiasing methods. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 2 (Short Papers)*, pages 15–20, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. doi: 10.18653/v1/N18-2003. URL https://aclanthology.org/N18-2003.

[56] X. Zhou, Y. Nie, H. Tan, and M. Bansal. The Curse of Performance Instability in Analysis Datasets: Consequences, Source, and Suggestions. *arXiv:2004.13606 [cs]*, Nov. 2020. URL http://arxiv.org/abs/2004.13606. arXiv: 2004.13606.

A Linguistic Acceptability

The Corpus of Linguistic Acceptability [CoLA; 48] is a set of acceptable and unacceptable English sentences collected from the linguistics literature. Linguistics has a longstanding practice of studying minimal changes that render sentences ungrammatical; one CoLA example is the pair of sentences "Betsy buttered the toast" (acceptable) and "Betsy buttered at the toast" (unacceptable).

CoLA includes an ID val/test set, where the examples are taken from the same linguistics papers that the training set uses. However, it also includes an OOD diagnostic val/test set. The diagnostic sets are taken from a different set of linguistics papers, so in order for a model to perform well on CoLA-OOD, it must transfer a general ability to recognize unacceptable English sentences, rather than simply learning the set of acceptability rules described in the ID sources.

A.1 Experimental details

We found that default settings on the HuggingFace [51] training script[7] resulted in more pronounced barriers between models, compared to the Google script we used for NLI and QQP. Because our goal is to study the relationship between barriers and generalization, we therefore chose to use Huggingface for our CoLA analysis. Like in our other experiments, we kept the default hyperparameters, which differ slightly from the Google script. The CoLA models were trained for 6 epochs with a learning rate of 2×10^{-5}, a batch size of 32 samples, and 0 weight decay. Both scripts use the AdamW[28] optimizer with a linear learning rate decay schedule and no warm-up.

A.2 Clustering

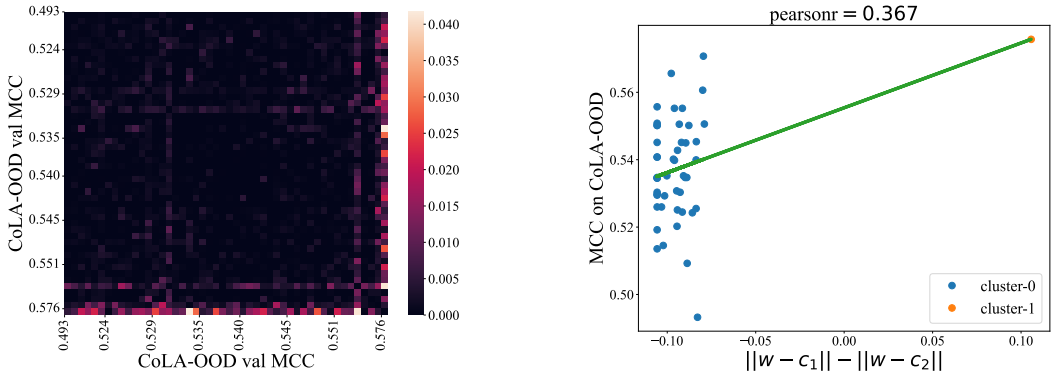


Figure 8: (a) $\mathfrak{CG}$ heatmap on CoLA models, sorted by OOD validation. (b) A scatter-plot of cluster membership versus performance on CoLA-OOD dev set.

In CoLA, there are very few barriers between finetuned models. A single model out of the 48 finetuned accounted for all substantial interpolation convexity gaps (Fig. 8(a)), thus forming its own one-point cluster when using $\mathfrak{CG}$ as a distance metric for spectral clustering. This outlier model outperformed all others on OOD generalization (Fig. 8(b)), suggesting that CoLA is another task where models with different generalization behavior are disconnected.

B Theoretical result on convexity gaps

Theorem 1. *An ϵ-basin will have $\mathfrak{CG}(w_1, w_2) \leq \epsilon$ for every pair of models w_1, w_2 on its surface.*

Proof. Recall the definition of convexity gap as the maximum value of the barrier height of any segment θ_1, θ_2 along the interpolation between w_1 and w_2 from Equation 4:

$$\mathfrak{CG}(w_1, w_2) = \sup_{\gamma,\beta} \mathfrak{BH}(\gamma w_1 + (1-\gamma)w_2, \beta w_1 + (1-\beta)w_2) \qquad \gamma 1, \beta \in [0, 1] \tag{5}$$

[7] `https://github.com/huggingface/transformers/blob/main/examples/flax/text-classification/run_flax_glue.py`

As we are in an ϵ-basin, the defining inequality from Equation 2 holds $\forall \theta_1, \theta_2 \in \epsilon-$convex basin :

$$\mathcal{L}(\sum_{k=1}^{n} \alpha_k \theta_k) \leq \epsilon + \sum_{k=1}^{n} \alpha_k \mathcal{L}(\theta_k) \tag{6}$$

Applying this for $n = 2$:

$$\mathcal{L}(\alpha_1 \theta_1 + (1 - \alpha_1)\theta_2) - (\alpha_1 \mathcal{L}(\theta_1) + (1 - \alpha_1)\mathcal{L}(\theta_2)) \leq \epsilon \qquad \forall \theta_1, \theta_2 \in \epsilon\text{-basin} \tag{7}$$

Hence the supremum of the quantity on LHS is also $\leq \epsilon$. Seeing the definition of $\mathfrak{BH}$ from Equation 3, we immediately see that:

$$\mathfrak{BH}(\theta_1, \theta_2) \leq \epsilon \qquad \forall \theta_1, \theta_2 \in \epsilon\text{-basin} \tag{8}$$

As $\mathfrak{CG}$ is the supremum of $\mathfrak{BH}$ over elements within the $\epsilon-$convex basin only, we have:

$$\mathfrak{CG}(w_1, w_2) \leq \epsilon \qquad \forall w_1, w_2 \in \epsilon\text{-basin} \tag{9}$$

□

C HANS performance on subsequence heuristics

Models that perform poorly on subsequence heuristics tend to fall in the bag-of-words basin. However, the clusters are less pronounced than for either constituent or lexical overlap heuristics (Fig. 9).

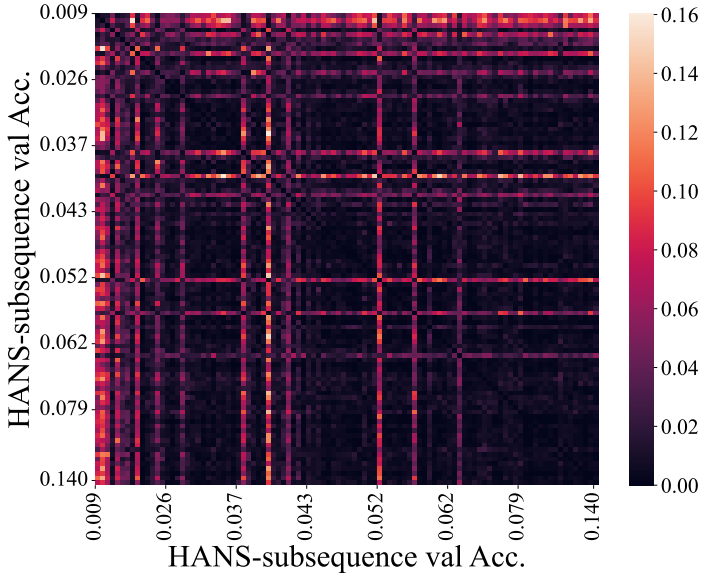


Figure 9: $\mathfrak{CG}$ for model pairs, sorted by increasing performance on HANS-subsequence.

D Flexible mode connectivity

While our results strongly suggest that linear mode connectivity breaks down between heuristic and generalizing models, they should nonetheless be connected, at least by a nonlinear path of constant training loss [6, 1]. We consider the behavior of HANS generalization along this path, which maintains near-constant ID loss. These paths were easy to identify with Riemannian curves and segmented lines [10], and exhibited poor HANS-LO generalization the further along the path the parameters fall, as shown in Fig. 10.

The results here show an inversion of the pattern exhibited on HANS-LO loss in the linear interpolation case: whereas Fig. 1(a) shows heuristic-to-generalizing connections to exhibit peaks in ID loss but no corresponding spikes in OOD loss, here we find constant-loss paths which exhibit peaks in HANS-LO loss. It appears that gradient descent finds generalizing models in spite of the fact that they exist in a constant-loss multidimensional volume which exhibits an apparent wealth of highly heuristic models.

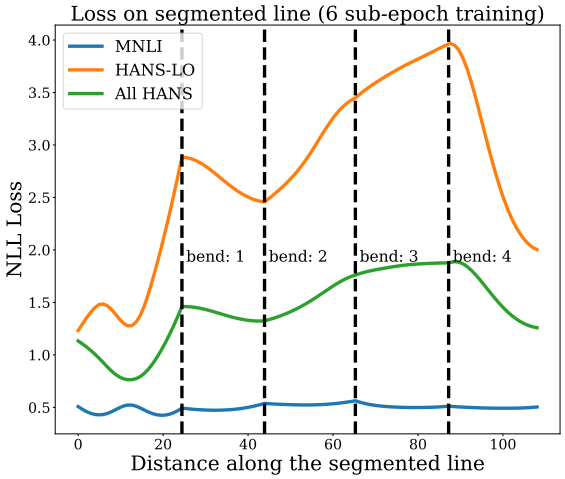


Figure 10: Loss along the segmented curve (after 6 subepochs) between a heuristic and generalizing NLI model. 1 sub-epoch is equivalent to $\sim$ 120 training steps with a batch size of 128, and a learning rate of 8×10^{-5}

E Correlations between performance on HANS subsets

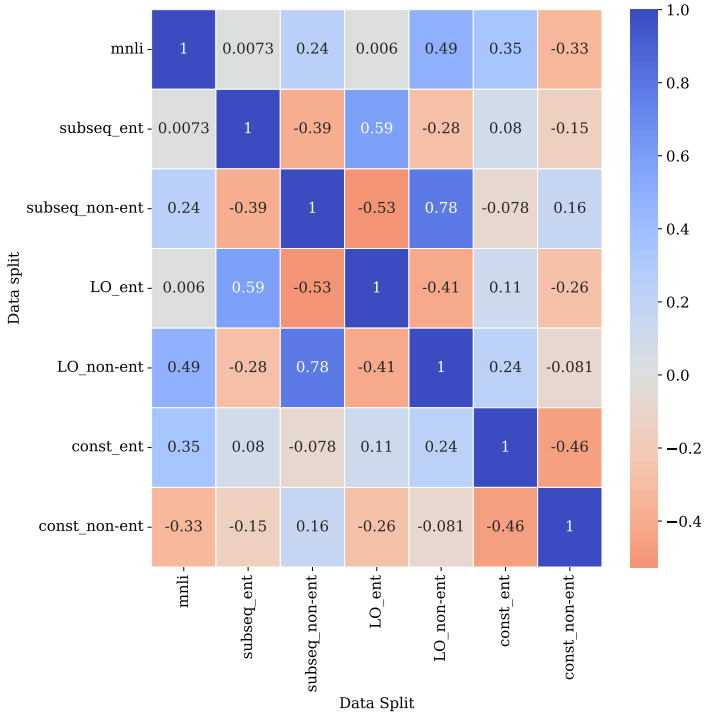


Figure 11: Correlation between accuracy on various ID and OOD diagnostic NLI test sets. MNLI refers to the MNLI validation set. Other sets were taken from HANS validation sets for subsequence, constituent, and lexical overlap heuristics. Both entailing and non-entailing subsets are included.

As provided in Fig. 11, performance on HANS-LO non-entailing, which reflects an aversion to the lexical overlap heuristic, is positively correlated with ID accuracy. However, in the case of HANS-constituent a model that successfully averts the constituent overlap heuristic is likely to perform poorly ID and on most diagnostic sets. We therefore conjecture that models which avoid

constituent overlap heuristics are likely failing to acquire syntax, rather than effectively acquiring better generalization strategies.

Performance on the subsequence heuristic diagnostic set is positively correlated with performance on both LO and constituent heuristic diagnostic sets, as it includes aspects of both other heuristics. For every diagnostic set heuristic, performance on the non-entailing subset (i.e., examples that violate the heuristic) is negatively correlated with performance on the entailing subset (i.e., examples that adhere to the heuristic).

F Experimental details

F.1 Finetuning the models

The QQP models were trained for 3 epochs, with a learning rate of 2×10^{-5}, a batch size of 32 samples and a weight decay of 0.01 from the `bert-base-uncased`[8] pre-trained checkpoint using the google script.[9] This script uses the AdamW [28] optimizer with a warm-up during the first 10% of the training followed by a linear schedule for weight decay. Because these hyperparameters are the recommended defaults for BERT training, they were also used by McCoy et al. [31] to train the MNLI model set we analyze.

F.2 Interpolations

In order to evaluate the interpolation between a pair of models, we sample a random set of 512 samples from the target dataset and then test each of the interpolated models on this same sample. For an interpolation between two fine-tuned models θ_1 and θ_2, we evaluate the models with interpolation coefficients α at intervals of $\frac{1}{10}$ on the line joining θ_1 and θ_2. The resulting plots are shown in Fig. 1(a) and Fig. 1(b). We calculate the convexity gap $\mathfrak{CG}(\theta_1, \theta_2)$ is using the loss values of these interpolated models.

In order to gauge the effect of number of samples, we repeated interpolation experiments on MNLI using only 256 samples (instead of 512). Using a smaller sample reduced the magnitude of correlation coefficient in Fig. 3(a) to $\rho = -0.65$, but the cluster assignments remain the same.

F.3 Model Evaluations

When evaluating the performance of a finetuned model (e.g., to sort the models on the linear connectivity heatmaps, or to select the top and bottom 5 models for Fig. 1(a) and Fig. 1(b)), we use the whole dataset available. These include the entirety of:

- All non-entailing samples of a given heuristic from the `test` split of HANS.
- The `dev_and_test` split of PAWS-QQP.
- The entire QQP and MNLI validation sets.

As Feather-BERTs have been finetuned on MNLI, to evaluate them on HANS tasks, we would need to convert the logits for the three MNLI classes (`entailment`, `neutral`, `contradiction`) to logits for the two HANS classes(`entailing` and `non-entailing`). We convert these labels by taking the logit for `non-entailing` class as the maximum of the logits of the MNLI `contradiction` and `neutral` classes, and taking the logit for `entailing` class of HANS as the logit of `entailing` class of MNLI.

F.4 Computing sharpness

For calculating ϵ-sharpness of a model, we use the definition of Keskar et al. [21]:

$$\phi_{x,f}(\epsilon, A) := \frac{\max_{y \in \mathcal{C}_\epsilon}(f(x + Ay)) - f(x)}{1 + f(x)} \times 100 \tag{10}$$

[8] `https://storage.googleapis.com/bert_models/2018_10_18/uncased_L-12_H-768_A-12.zip`

[9] `https://github.com/google-research/bert`

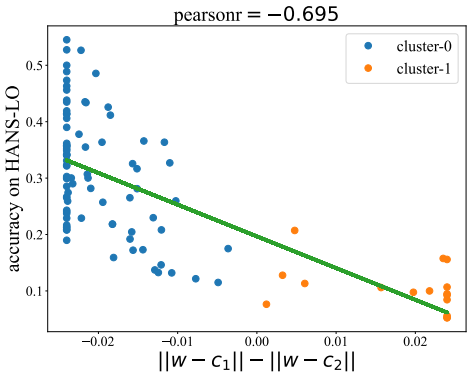
where f is the loss function, $x \in \mathbb{R}^n$ are the original parameters of the model, $A \in \mathbb{R}^{n \times p}$ is a matrix that restricts the calculation of $\epsilon-$sharpness to a subspace of the parameter space, and

$$\mathcal{C}_\epsilon = \{z \in \mathbb{R}^p : -\epsilon(|(A^+x)_i| + 1) \leq z_i \leq \epsilon(|(A^+x)_i| + 1) \forall i \in [p]\} \tag{11}$$

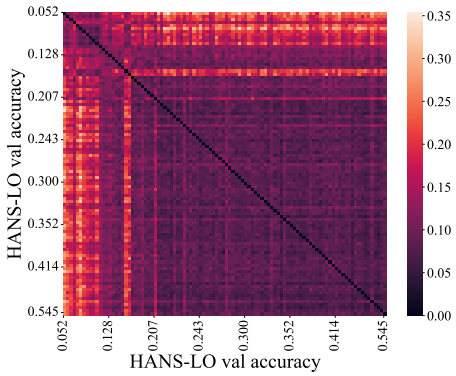
Following the defaults in Keskar et al. [21], we take A as $\mathbb{I}_{n \times n}$. Using 32768 samples from MNLI `validation_matched`, we evaluate the initial loss $f(x)$. We use the same samples to compute the maximum loss $\max_{y \in \mathcal{C}_\epsilon} f(x + Ay)$, using an SGD optimizer with a learning rate of 8×10^{-5}, a batch size of 32, and accumulating gradient over 4 batches. In order to compute maximum loss, we perform a total of 8192 gradient updates, each followed by clamping of the weights within $\mathcal{C}_\epsilon$. We set ϵ to 1×10^{-5}.

G Training loss

Even without any validation or test set to analyze, we can use the training set alone to make predictions about a model's generalization strategies. Fig. 12 shows a strong relation between training loss convexity gaps and performance on HANS-LO. Future work can develop methods for predicting generalization early based entirely on training behavior.



(a) Clusters are based on convexity gaps in the MNLI training loss surface. Best squares fit shown.



(b) Color indicates convexity gap. Models on axes are sorted according to OOD performance.

Figure 12: Relationship between $\mathfrak{CG}$ on the MNLI *training loss surface* and accuracy on the HANS-LO diagnostic set.

H In-Domain Generalization

train	test	f	μ_{C_1}	σ_{C_1}	μ_{C_2}	σ_{C_2}	$\frac{\mu_{C_1} - \mu_{C_2}}{\sigma_{C_1}}$	$\frac{\mu_{C_1} - \max_{w \in C_2} f(w)}{\sigma_{C_1}}$
MNLI	MNLI	acc	0.844	0.002	0.839	0.002	2.50	1.00
MNLI	HANS-LO	acc	0.301	0.110	0.096	0.033	1.86	1.30
QQP	QQP	F1	0.879	0.001	0.872	0.002	7.00	5.00
QQP	PAWS-QQP	F1	0.436	0.005	0.424	0.004	2.40	0.60
CoLA	CoLA ID	MCC	0.601	0.017	0.600	N/A	0.06	0.06
CoLA	CoLA OOD	MCC	0.537	0.015	0.576	N/A	-2.60	-2.60

Table 1: Cluster mean, standard deviation, and distribution overlap for all tasks on both ID and OOD performance under metric f. C_1 is the larger cluster on each task. On CoLA, C_2 contains only a single model.

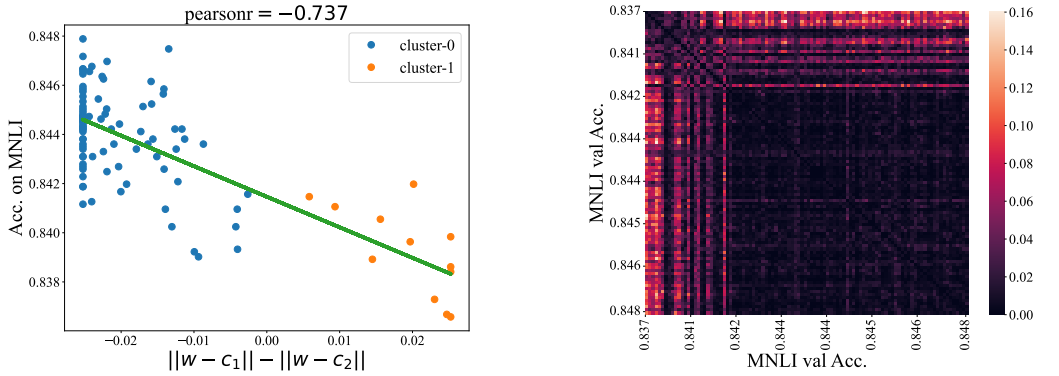
Let us consider the skeptical position that basin membership does not directly relate to OOD generalization strategies, but instead OOD performance is mediating the effect of general model quality in-domain. While ID (on MNLI) and OOD (on HANS) performance are certainly positively

correlated (Appendix E), we find that all 12 of the models in the syntax-unaware cluster fall at least 1.3 standard deviations below the mean of the syntax-aware cluster's performance on HANS-LO (Table 1), and only 1 standard deviation below the mean on MNLI validation. The partitions are also less clear in the ID metrics (Fig. 13) compared to lexical overlap. Therefore, considered as a binary class, basin membership is more predictive of OOD behavior than of ID on NLI. Furthermore, the fact that performance on HANS-constituent is *negatively* correlated with ID performance (Appendix E) makes it clear that the underlying difference between these clusters is syntax-awareness, not overall model quality.

CoLA presents even stronger evidence for interpreting basin membership as related to OOD generalization over ID. The single outlier model falling outside of the main basin is a full 2.6 standard deviations outside the mean on OOD performance, but only 0.6 on ID performance.

However, we should not dismiss the relationship between ID accuracy and cluster membership. Compared to the diagnostic set PAWS, the ID validation set is more predictive of $\mathfrak{CG}$ on QQP (Fig. 14). In both MNLI and QQP, the relationship between ID and centroid distance also seems more linear than OOD.

An important caveat is the fact that ID evaluation and ID convexity gap measurements are performed on the same validation set, because the test sets are not public for these datasets. Therefore, ID performance is exposed during the clustering process, casting some doubt on alignment between the clusters and ID performance.



(a) Clusters are based on convexity gaps in the MNLI validation loss surface. Best squares fit shown.

(b) Color indicates convexity gap. Models on axes are sorted according to ID performance.

Figure 13: Relationship between $\mathfrak{CG}$ on the MNLI in-domain validation loss surface and in-domain validation accuracy.

I Alternative barrier measurements

In Fig. 15 we show NLI model clustering using the original barrier height metric (Equation 3) from Entezari et al. [7]. We can see that the Pearson's correlation coefficient is -0.49, which is far lower in magnitude than correlation in the $\mathfrak{CG}$-metric space.

We also show the results for another metric, viz. area under the interpolation curve (AUC), in Fig. 16. Although this shows the same Pearson's correlation coefficient, the clusters are much less crisp in the heatmap. In order to compute AUC, we first subtract the lowest point on the curve from all points, and then compute the area under the shifted curve.

Finally, we consider the effect of Euclidean distance (Fig. 17). The clustering effect is extremely strong and predictive of generalization behavior. It is worth considering the possibility that basins trap the models in a way that forces them to have larger Euclidean distances, but those Euclidean distances are the property that most determines the generalization strategy of the model.

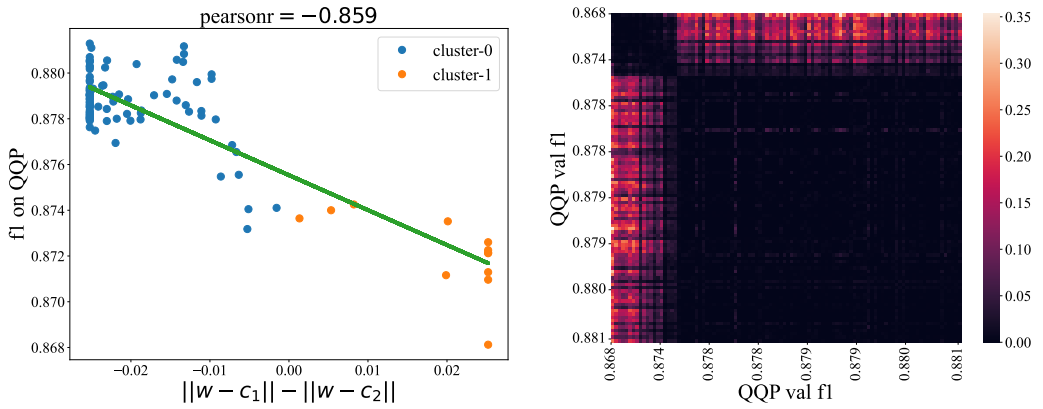


Figure 14: Relationship between $\mathfrak{CG}$ on the QQP in-domain validation loss surface and in-domain validation accuracy.

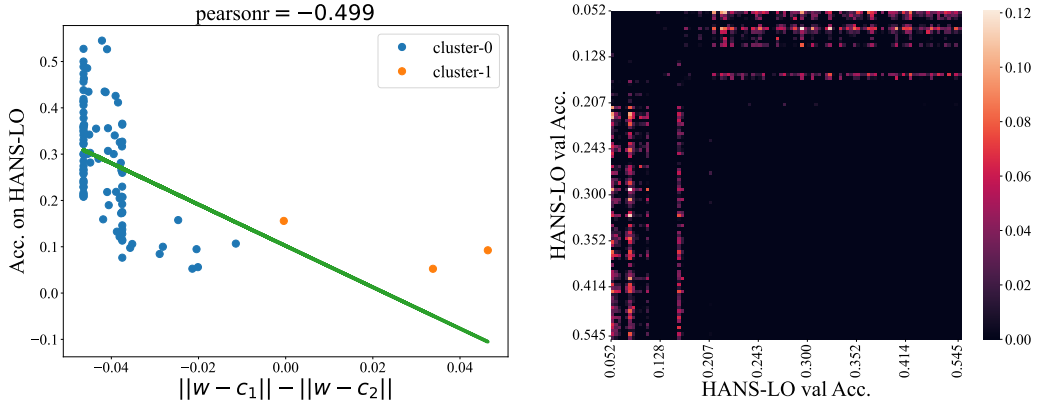


Figure 15: Relationship between $\mathfrak{BH}$ on the MNLI validation loss surface and HANS-LO accuracy.

J Code and Models

We provide an anonymized version of our code at: `https://github.com/aNOnWhyMooS/connectivity` and all our fine-tuned models and evaluation as well as interpolation logs are released at: `https://huggingface.co/connectivity`.

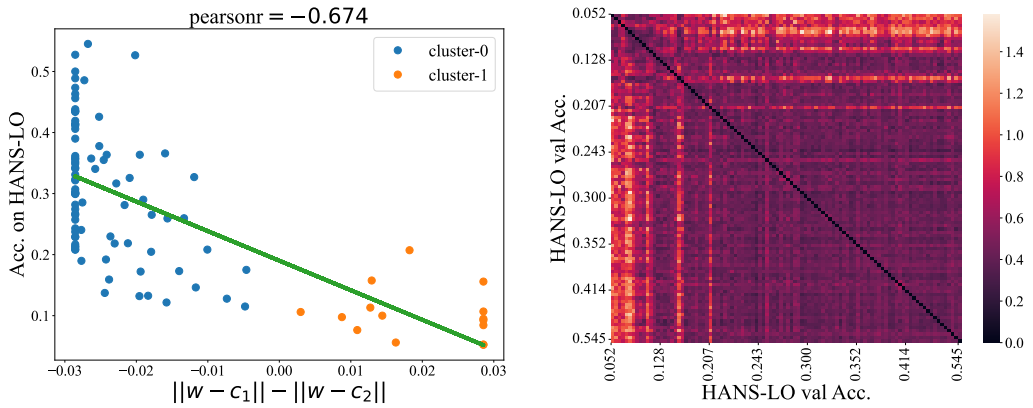


Figure 16: Relationship between AUC on the MNLI validation loss surface and HANS-LO accuracy.

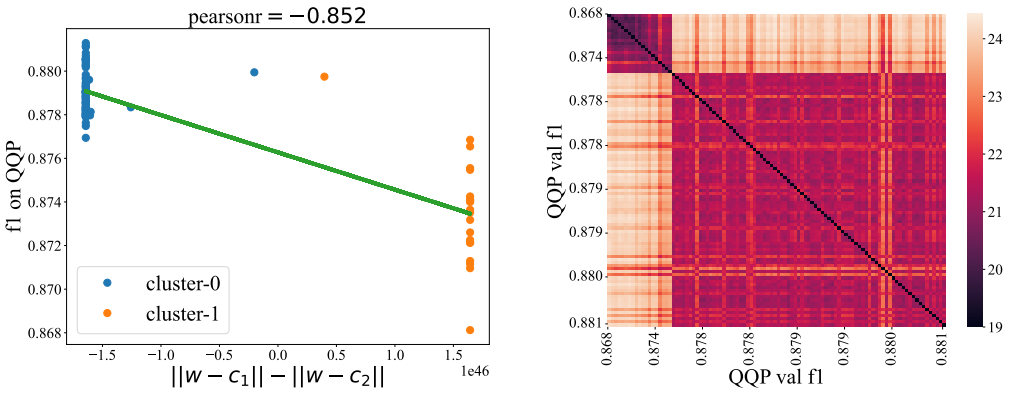


Figure 17: Relationship between Euclidean distance on the QQP validation loss surface and PAWS-QQP accuracy.

LOW THD GRID-CONNECTED INVERTER FOR TWO PV ARRAY

Vinay, Parag Agrawal, Neeraj Kumar
Delhi Technological University, Delhi, India

Abstract- Grid-connected inverters based on solar PV cells are becoming increasingly common. The PV can sometimes be subjected to a certain environmental situation, such as varying irradiance or temperature. In this case, present developments in grid-connected PV inverters do not have many reliable yields or reliability. In this article, an improved control framework system is intended for two PVs that are mismatched in various ways, resulting in a highly effective grid-connected solar inverter with THD under 1% for both voltage and current. The regulator employs a control scheme comprised of an ANFIS-based control system that works in tandem with the incremental conductance technique of MPPT. The outcomes are simulated and validated using MATLAB Simulink.

Keywords- *Solar PV inverter, ANFIS controller, 2 PV array, Power system*

Introduction:

Essentially, a photovoltaic array is a connected group of solar modules, one of which is seen on the right in the illustration. Each photovoltaic (PV) module is made up of several PV cells that are coupled together. Solar energy is converted into direct current electricity by solar cells. Solar photovoltaic (PV) modules are frequently referred to as solar panels, although the word is more appropriately used for solar-thermal water or air heating panels. Photographic modules separate themselves from solar cells in that they are easily scaled and packed in weather-resistant housings for ease of installation and deployment in a variety of applications including residential, commercial, and industrial settings. Photovoltaics is the term used to describe the use and research of solar devices.

Land, water, pollution, dangerous chemicals, noise, and visual pollution are all examples of the negative environmental implications of PV systems. PV system design trends for the future will emphasize enhanced design, long-term sustainability, and recycling. Incentives and research to fix the gaps may serve as a solid foundation for future laws and regulations. Carbon emissions, hazardous waste, unsustainable mining techniques, and habitat destruction are all associated with the manufacture of photovoltaic panels. Environmental concerns, along with solar's limited ability to produce enough electricity to power the grid, should be taken into account by both citizens and decision-makers when making solar energy decisions.

In this work, a 2 PV array of different environmental conditions is made using an ANFIS controller. The growing popularity of renewable energy sources over the past several decades has gained speed as a result of the continued shortage of fossil fuel resources. This has also increased the importance of electrical energy in the world today. In this context, the photovoltaic (PV) business has been developing at an alarmingly fast pace over the last few years. Photovoltaic (PV) systems have the potential to store all of the world's electrical output. Starting late, the enthusiasm for unblemished, reasonable, and supportable force sources has extended dependably since the oil-based commodities have made normal issues. Accordingly, elective vitality sources are settled for the issues, for instance, sunlight-based vitality and force modules. Among the collection of practical force source resources accessible, sun-oriented vitality has exceptional thought taking into account its wealth and defilement free change to power through the photovoltaic (PV) process. Extending eagerness for PV frameworks, for instance, MPPT (Maximum Power Point Tracking) needs a continuous number of PV displays foundation & force electronic circuits to enhance adequacy, safety level, resolute quality & quality of force. In this paper, a PV inverter with mismatched environmental conditions is made.

Literature Review

[1] S Dutta. *et al.*, It is proposed in this paper that a single-phase grid-connected transformerless photovoltaic (PV) inverter, which can operate in either buck or boost mode and can extract maximum power simultaneously from two serially connected subarrays while each subarray is subjected to a different environmental condition, be used to extract maximum power from two serially connected subarrays. Because the inverter may function in either buck or boost mode, depending on the application, the restriction on the minimum number of serially linked solar PV modules that must be used to construct a subarray is substantially decreased. Because of this, when each subarray is subjected to a distinct set of environmental circumstances, the power yield from each subarray rises. It is intended that the inverter's topological configuration and its control approach be such that the high-frequency components are not present in the common-mode voltage, hence keeping the amplitude of the leakage current associated with the PV arrays within a given range. A high level of operational

efficiency is also obtained across the whole working range. After doing a thorough investigation of the system, which ultimately leads to the construction of a mathematical model of the system, It is determined whether or not the project is viable by carrying out comprehensive simulation studies. It is necessary to create a 1.5 kW laboratory prototype, and extensive experimental tests must be carried out to confirm the correctness of the concept.

[2] Iqbal *et al.*, Grid Link Inverter and MPPT-A Review, this article gives an outline of the system Link Inverter (GTI), PV Inverter, Inverter Topology, and MPPT System (Also Known as Maximum Power Point Tracking) and Its sorts, for example, Steady Voltage, Fixed Duty Cycle, Modified P&O, and P&O Perturb and Observe, IC Incremental Conductance and Modified IC, Ripple Link and System Wavering Techniques give subtleties of how these strategies have been executed.

[3] Hong Li *et al.*, "Control of Cascaded Inverters with MPPT for Grid-Connected Photovoltaic Generators" Alone stage course inverter for cross-section related PV framework with stage move control is presented in this paper.

[4] Mohd Alam *et al.*, This paper deals with a switched lift inverter (SBI) applicable for the sun-powered photovoltaic framework (PV) interfaced scaled downscale network. SBI is a forlorn stage control converter that can gracefully dc and ac stacks for microgrid applications at the same time. SBI's creation voltage can likewise be more striking or not as much as its voltage dc information. This converter likewise permits the inverter legs to be shot without hurting the converter. The rule for the activity of the SBI as a force forming unit in a sunlight-based PV interfaced microgrid is explained in detail in this paper. To confirm the proposed converter and its control framework interfaced with sunlight-based photovoltaics, MATLAB-based re-enactments are led. Besides, in the propagation, MPPT is connected to improve its sensibility in the activity of little scope, interfaced sunlight-based photovoltaic grid in different viewpoints.

[5] Hors *et al.*, Low Power REGS is one of the basic employments of electronic force converters, a switched Capacitor Multilevel Inverter for Grid-Connected PV Systems with MPPT ' Ability and Reduced Components. This paper suggests an extra application in REGS for a solitary stage lattice-connected switched condenser inverter. According to the control procedure & Seri equal trade of semiconductors, it has controllable yield voltage wealth with low full musical mutilation in examination and correlation topologies & has no unpredictable figuring to modify the voltage of the condensers independently

[6] Manel Hlaili, *et al.*, This single-composed topology is more compact than standard topology, it has been chosen taking into account the truth that with the best control technique conceivable. Partnering the PV boards with the vitality network is reasonable. To interface the inverter to the cross-section, a stage bolted circle control is utilized. At that stage, the inverter's data voltage is overseen by a recommended DC-Link voltage order. Regardless of the truth that an MPPT estimation was utilized to advance the evacuation of power and the productivity of the plan. Inverter Output Current screen to create a little Total Harmonic Distortion (THD) input (present installed in the force arrange) acted in a DSP. Propagation and test results confirm the best possible working of the recommended plot, even with sun-powered radiation vacillations.

[7] Shuai Zhang *et al.*, To interface the inverter to the cross-segment, a phase darted circle control is used. At that stage, the inverter's information voltage is supervised by a suggested DC-Link voltage request. Despite the reality that an MPPT assessment was used to propel the departure of force and the profitability of the arrangement. Inverter Output Current screen to make a little Total Harmonic Distortion (THD) input (present introduced in the power organize) acted in a DSP. Engendering and test outcomes affirm the most ideal working of the suggested plot, even with sun-controlled radiation instabilities.

[8] P. Chinna D. *et al.*, "Overall MPPT of Grid related Solar PV Inverter under Partially Shaded Condition "the control delivered by a sunlight-based photovoltaic (SPV) framework depends on the sun-powered radiation level and cell temperature. A sensible greatest force point tracking (MPPT) count is required to extract the most extreme force from the SPY. Regardless, at a midway concealing condition (PSC) the customary MPPT systems may incite the activity at a local most extreme force point (LMPP) as opposed to the overall greatest force point (GMPP). This paper used a GMPPT count to extract the greatest force from the sun-based PV framework, and a DC/AC inverter is used to change over-extracted sunlight-based DC control into AC capacity to mix into the grid. A three-leg sharp force module (IPM) is used for the execution of both the DC-DC converter and DC-AC inverter.

[9] Ahmed Sony Kamal Chowdhury *et al.*, This article proposes a topology for single-stage two-stage cross-section related sun powered photovoltaic (PV) inverter for private applications. Our proposed organized related force converter involves a switch-mode DC-DC help converter and an H-interface inverter. The trading procedure of the proposed inverter contains a blend of sinusoidal heartbeat width change (SPWM) and square wave close by network synchronization conditions. The execution of the proposed inverter is mirrored under structure-related circumstances by methods for PSIM. Besides, the sharp PV module framework is actualized using a direct most extreme force point tracking (MPPT) methodology utilizing power balance is moreover used to grow the capability of the framework,

[10] Marcio Mendes Casaro *et al.*, This article presents a plan associated with PV grid in an arrangement that is joined and made with a twofold stage three-phase inverter. In this topology, generally speaking, the most elevated vitality point observing (MPPT) is happened by a DC to DC converter and the cross-section current is constrained by an inverter. The three-phase plot total converter (SRC3) is chosen thankfulness for the helpful conditions it shows for the DC-DC agreement.On any occasion, the achievement of MPPT is lacking because of its capacity to depend unambiguously on the actualized cutoff time and trade repeat.

[11] LISheng-qing *et al.*, Photovoltaic grid-related inverter framework has the upsides of clear topology and insignificant exertion. Since the yield force of photovoltaic contraptions is a nonlinear capacity of the external condition load, to help the execution of photovoltaic devices, Maximum Power Point Tracking (MPPT) control should change the obligation cycle disrupting impact reliant on created by photovoltaic devices. Even though MPPT control with using customary aggravation and recognition strategy has the advantages of clear in structure and easy to execute the hardware circuit, anyway the control is dubious and generally poor.

[12] T. Kerekes *et al.*, "MPPT figuring for Voltage Controlled PV Inverters" This paper shows an original thought for an MPPT that can be used if there ought to be an event of a voltage-controlled framework related PV inverters Dependent upon the DC interface capacitor, this force vacillation could be worked to track MPP of PV show, Force movements are nearly nothing using the information that at MPP.

[13] Sonam S. Katkamwar *et al.*, "Fell H-Bridge Multilevel PV Inverter with MPPT For Grid Connected Application" a three-phase five measurement fell H-interface PV (photovoltaic) inverter with Maximum Power Point Tracking (MPPT) for structure-related application. Interconnection of sun based & structure is a maximum mentioning point nowadays. For fell, H-associate stunned inverter with PV group as free DC source Standard inspiration driving this paper is to make control framework. Estimated topology upgrades capability & flexibility of PV framework. Used individual Maximum Power Point Tracking (MPPT) strategy for each sun-based board to extract the most extreme force from the sunlight-based board and used individual control plan of each dc associate voltage. For three-phase cross-section-related applications, due to PV clutters it is inconsistent given force. To deal with that issue using a control scheme with modification pay. Entertainment result is addressing check control procedure. By using MATLAB/Simulink Re-enactment show has been created.

[14] Jyothirmayi. C. J. *et al.*, "Step Modulated Multilevel Inverter Incorporated Upon ANFIS based Intelligent PV MPPT" Building Integrated sunlight-based cells has now transformed into a rising trend in the domain of current drives and vitality change. Consolidated PV board has execution adversities on account of rapid assortment in sun-powered light power and temperature of the sun-oriented board. This necessity on the PV side is regulated by a suitable Adaptive neuro-fuzzy [ANFIS] based astute Maximum Power Point Tracking (MPPT) control computation, Framework moreover combines a force forming unit (Inverter)

[15] CC Marouchos *et al.*, The new dc-to-ac converter topology relying upon the turned condenser framework was given late. The fruitful instances of low consonant substance, raised proficiency, MPPT facility, and voltage adventure up highlights make it a decent open door for maintainable vitality sources associated Additionally, in this archive, the musical material is confined by showing different present modulators, each associated with a PV board. With the structure. It is additionally demonstrated that the circuit likewise screens the Maximum Power Point of the photovoltaic board's yield quality with developing sun and adjacent temperature illumination. Also, in this report, the musical material is confined by showing different present modulators, each associated with a PV board. The activity is done with broad PSIM multiplications

Implementation

Currently, the execution used in the previous chapter has been modified for the proposition of the current job from the foundation work [1]. Rather than using just the IC MPPT conductance technique, an ANFIS-based control regulator is used in this work for greater precision and lower distortions. In Figure 1, the circuit outline is taken as is to allow an inspection for effects, and ANFIS is used to boost this circuit.

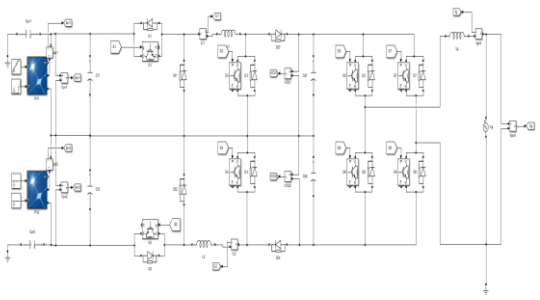


Figure 1: Final Model for Two PV array Inverter

Figures 2 and 3 show the ANFIS regulator, and in the subsystem the power is measured, resulting in a change in power that is incorrect and a change in power noise as two sources of information that are multiplexed and taken into the input of the ANFIS regulator in an attempt to boost the error that was generated by using PV arrays in mismatched environmental conditions.

This yield has a dimension that increases the distortion proportion for both current and voltage outputs.

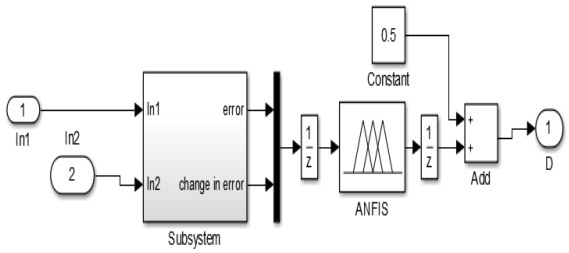


Figure 2: ANFIS Controller for Two PV inverters

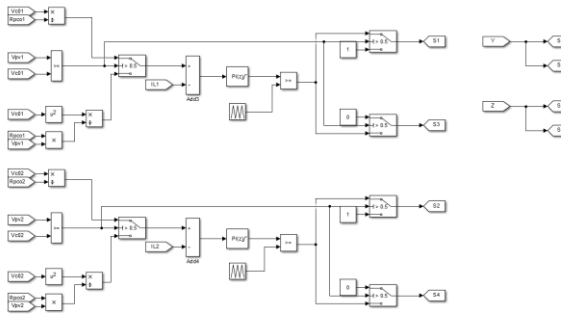


Figure 3: ANFIS Controller for Two PV inverters

Figure 4 shows how the subsystem is defined in terms of error and improvement in distortion yield from PV current and voltage. Since the PV exhibits are mismatched and can result in different yields and mismatches, two ANFIS regulators are used for each PV cluster.

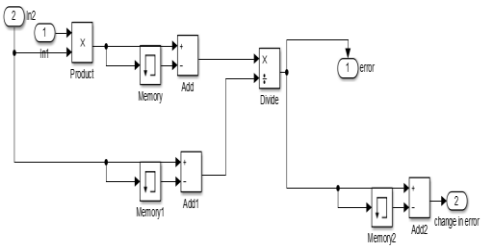


Figure 4: Subsystem for Power Error and Change in Error

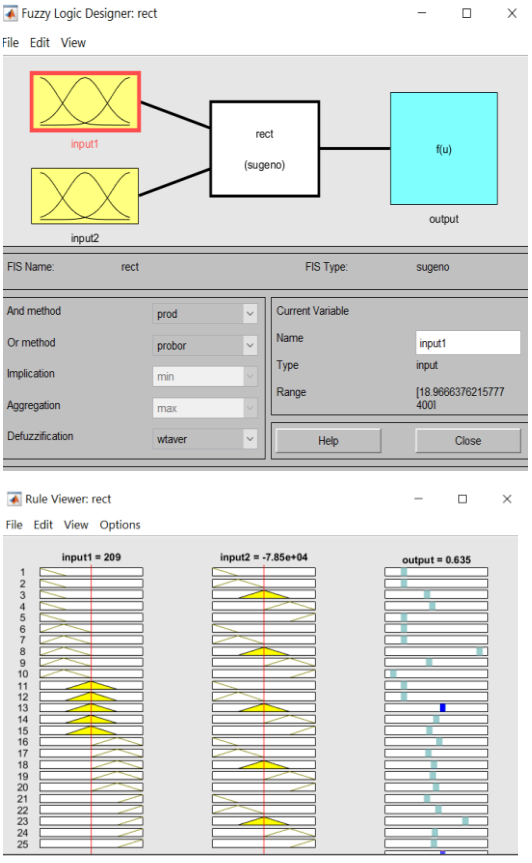
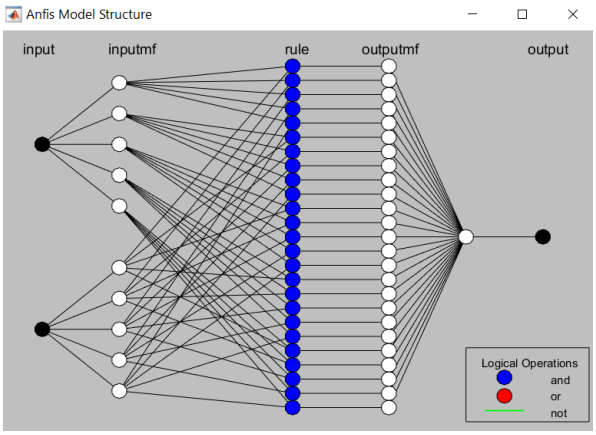


Figure 5: ANFIS Toolbox and rule view for Power Error and Change in Error and Output

Figure 5 depicts the ANFIS regulator instrument, which demonstrates the Sugeno requirements used by ANFIS regulators.



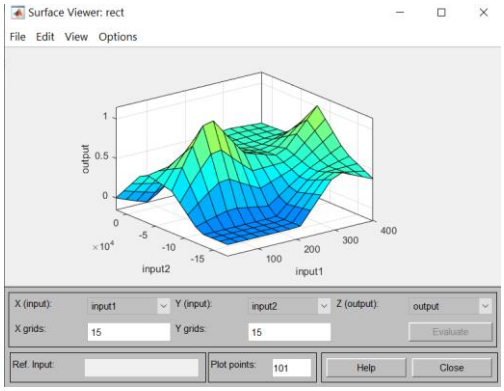


Figure 6: The Surface View and model structure of the Rules range

Figure 6 depicts a surface view graph of the rules described in two PV arrays ANFIS controllers.

Results

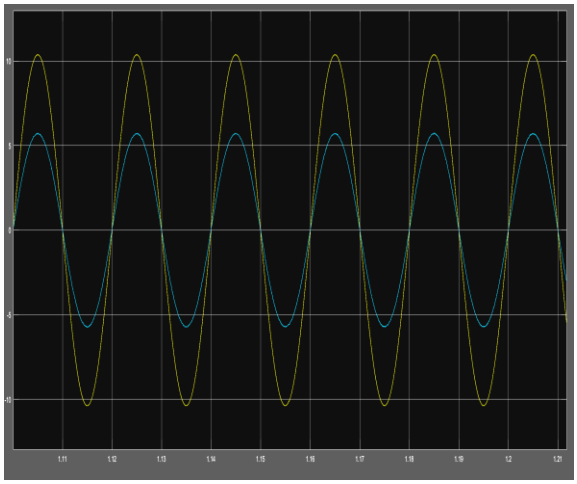


Figure 7: Output Current and Voltage of Inverter

The zoomed output of current and voltage in Figure 7 shows a sine wave, which is perfectly fine.

The final comparison is seen in table 1 below, based on the findings shown in the previous section. For planned work, the outcomes are improved.

Table 1: Comparison results for Two Inverters

	Existing	Proposed
THD Voltage	0.09%	0.05%
THD Current	2.12%	0.43%

In addition, the data are described graphically in contrast, as seen in figures 8 and 9 below.

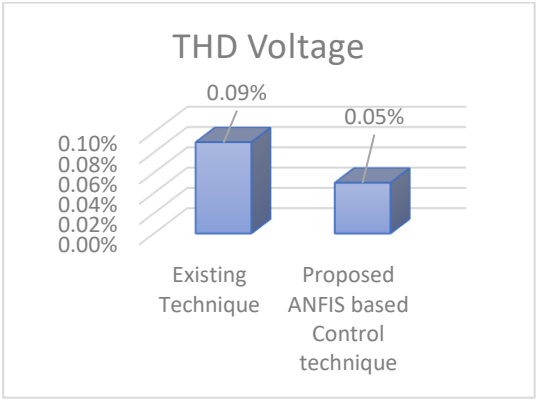


Figure 8: THD comparison for Voltage

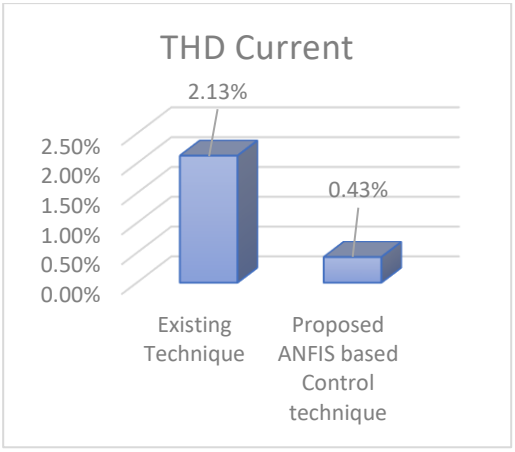


Figure 9: THD comparison for current

Figure 10 and Figure 11 show the THD Outputs for voltage and current.

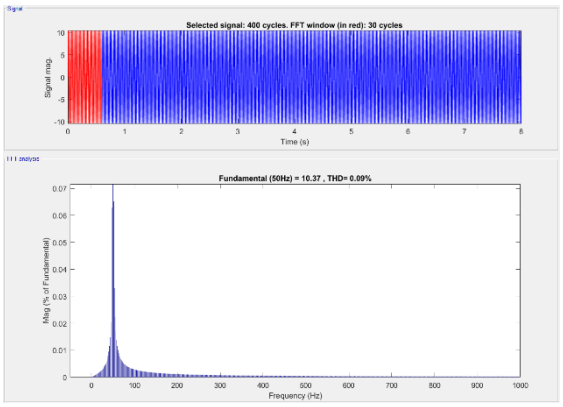


Figure 10: THD output for Voltage

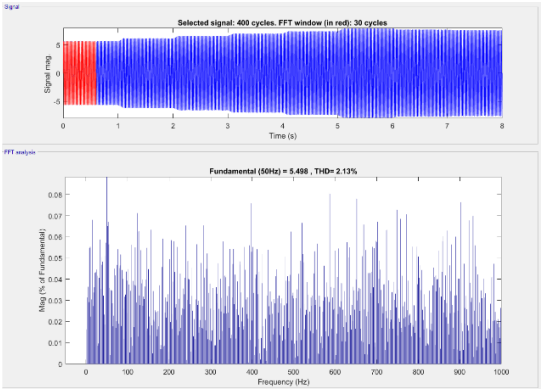


Figure 11: THD output for Current

Conclusion

The execution of two PV-based arrays in mismatched natural environmental conditions is actualized in this paper to form an inverter with fewer noises and distortions. The experiment is carried out efficiently, with the findings shown in MATLAB Simulink. The distortions are less than 0.5 percent, and the yields have an efficiency of more than 99.5 percent. The yield current and voltage waveforms demonstrate correct yield, and low variances for PV yields are monitored. The ANFIS regulator is the most accurate because it takes into account the problematic states of error detections and changes in power mismatch in MPPT.

References

[1] S. Dutta and K. Chatterjee, "A Buck and Boost Based Grid Connected PV Inverter Maximizing Power Yield From Two PV Arrays in Mismatched Environmental Conditions," in IEEE Transactions on Industrial Electronics, vol. 65, no. 7, pp. 5561-5571, July 2018, DOI: 10.1109/TIE.2017.2774768. [2] Iqbal, Zafar & Djokic, Sasa & Yanchenko, Sergey & Spelko, Aljaz & Papic, "Performance Evaluation of a PV Inverter with Active Filter Functionalities," *IEEE Transactions on Power Electronics.*, pp. 852-857, 2020.

[3] Hong Li, Jaber Abu-Quahog, "Control of Cascaded Inverters with MPPT for Grid-Connected Photovoltaic Generators," *IEEE Transactions on Power Electronics.*, Intelec(r) 2013 · 13. - 17. October 2018.

[4] Mohd Alam, Joydip Jana, Hiranmay Saha, "Switched Boost Inverter Applicable for Solar Photovoltaic System Based Micro-Grid," 2nd International Conference on Control, Instrumentation, Energy & Communication (CIEC), 2016.

[5] Hosse, Nasiri Avanaki, Reza Barzegarkhoo, Elyas Zamiri, "A Switched-Capacitor Multilevel Inverter for Grid-Connected PV Systems with MPPT Ability and Reduced Components," 9th Annual Power Electronics, Drives Systems, and Technologies Conference (PEDSTC), 2019.

[6] Manel Hlaili, Hfaiedh Mechergui, Henrique Gonçalves, Bruno Exposto, João L. Afonso, "Single Phase NPC Inverter Controller with Integrated MPPT for PV Grid Connection," STA'2016, Sousse, Tunisia, December 2016.

[7] Shuai Zhang, Weimin Wu, Housing Wang, Min, "Single-Stage MPPT Control Realization for Aalborg Inverter in Photovoltaic System," Photovoltaic Power Systems Program. Report IEA-PVPS T1-13: 200

[8] P. Chinna D. Goud, "Global MPPT of Grid-connected Solar PV Inverter under Partially Shaded Condition," *IEEE Trans. Ind. Electron.*, vol. 56, no. II, pp. 4374-4380, Nov. 2009.

[9] Ahmed Sony Kamal Chowdhury1, M. Abdur Razzak, "Single-Phase Grid-Connected Photovoltaic Inverter for Residential Application with Maximum Power Point Tracking," *IEEE Journal of photovoltaics.*, vol. 3, no. 1, pp. 500-508, Jan. 2013.

[10] Marcio Mendes Casaro, Denizar Cruz Martins, "Grid-Connected PV System Using A Three-Phase Modified Dual-Stage Inverter," *IEEE Transactions on Industrial Electronics.*, vol. 53, no. 4, pp. 1002-1016, August 2006.

[11] Lisheng-qing, Zhang Bin, Xutian-jun, Yang Jun, "A New MPPT Control Method of Photovoltaic Grid-connected Inverter System," Proceedings of the CSEE, vol. 32, pp. 149 -153, 2012.

[12] T. Kerekes, R. Teodorescu, M. Liserre R. Mastromauro, A. Dell'Aquila, "MPPT algorithm for Voltage Controlled PV Inverters," *IEEE Transactions on Industrial Electronics*, vol. 53, no. 4, August 2006.

[13] Sonam S. Katkamwar, V.R. Doifode, "Cascaded H-Bridge Multilevel PV Inverter with MPPT For Grid Connected Application," *IEEE Trans. Ind. application.*, vol 5, no. 2, pp.1722-1731 March/april 2015.

[14] Jyothirmayi C. J., Nasar. A., "Step Modulated Multilevel Inverter Incorporated Upon ANFIS based Intelligent PV MPPT," International Conference on Magnetics, Machines & Drives, 201

[15] C.C. Marouchos, "The Switched Capacitor Inverter as A MPPT In A Photovoltaic Application," IEEE 33rd Annual, pesc02, 2002.

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360238497

Manufacturing of Regeneratively Cooled Rocket Nozzles

Article *in* International Journal of Innovative Research in Science Engineering and Technology · April 2022

DOI: 10.15680/IJIRSET.2022.1104108

CITATIONS
0

READS
56

4 authors, including:



A. Ospanova
Al-Farabi Kazakh National University
248 PUBLICATIONS **593** CITATIONS

SEE PROFILE

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

MAY SCOPUS View project

Infomatics View project

All content following this page was uploaded by A. Ospanova on 28 April 2022.

The user has requested enhancement of the downloaded file.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

Manufacturing of Regeneratively Cooled Rocket Nozzles

Dr. A.K. Madan[1] , Prakash Narayan Budhaulia[2], Priyanshu Nailwal[3]

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

Department of Mechanical Engineering, Delhi Technological University, Delhi, India[2,3]

ABSTRACT: Liquid rocket engines components take complicated manufacturing processes because there is a need for nozzle cooling methodologies which need the construction of channels. Regeneratively- cooled (regen) or dump-cooled channel wall nozzles make use of an substantial area proportion outline to supply expansion of hot gasses plus active cooling of the hot wall to take care of acceptable structural perimeters. Large-scale component structures demand veritably thin walls for the channels and they need to survive extreme atmospheres handing challenging thermal and structural loads.
Additive manufacturing is the process of making an object by fabricating it one layer at a time. It's the alternative of subtractive manufacturing, during which an object is formed by cutting away at a solid block of fabric until the ultimate product is complete. Technically, additive manufacturing can confer with any process where a product is made by fabricating substance over, like moulding, but it normally refers to 3-D printing.

KEYWORDS: rocket, nozzle, additive manufacturing, 3-D printing

I. INTRODUCTION

The last decade has seen great developments within the space sector. Space and rocket science are not about transporting humans into space or exploring other planets, but also about doing investigation projects and transferring satellites into orbit for bettered remote sensing operations. This has increased the demands for launch vehicles and competition too because the sector is now expropriated by private players.
Rockets are extremely complicated and plethora of cash is invested into them for study and development. Manufacturing them could be a lot harder and more complicated than exploration work and if there are any failures, also the complete engine might need to be evaluated for the source of failure and further reconstructed for another test. This may be both costly and time- consuming which may be a problem everyone would want to omit. therefore, to unravel the toughest part of launching a vehicle, the manufacturing process is accelerated and this could be done using additive manufacturing or 3D printing.

In this paper, we have mentioned the manufacturing of channels of a regeneratively cooled nozzle and an entire 3D printing of the thrust chamber.
varied operations of additive manufacturing
* Aerospace material interest for AM
* AM for space operations
* AM for UAV operation
* Implicit challenges of AM in aerospace

As performance demands increase, the complexity also increases and traditional methodologies are complex. Using additive manufacturing, several operations are reduced to one process. Primary additive manufacturing methodologies that are studied are powder bed fusion (PBF) or selective laser melting (SLM) but these do not appear to be possible for giant scale production. consequently, alternative methodologies are used for big-scale nozzles and chambers including directed energy deposition (DED) processes.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

II. NOZZLE CHANNEL WALL

Liquid rocket engines building blocks demand complicated manufacturing processes because there is a need for nozzle cooling methodologies which necessitate the construction of channels. Regeneratively- cooled (regen) or dump- cooled channel wall nozzles make use of an outsized area proportion outline to produce expansion of hot gasses as well as active cooling of the hot wall to keep up satisfactory structural perimeters.
Large-scale building block structures demand very thin walls for the channels and since they've to survive extreme atmospheres delivering arduous thermal and structural loads

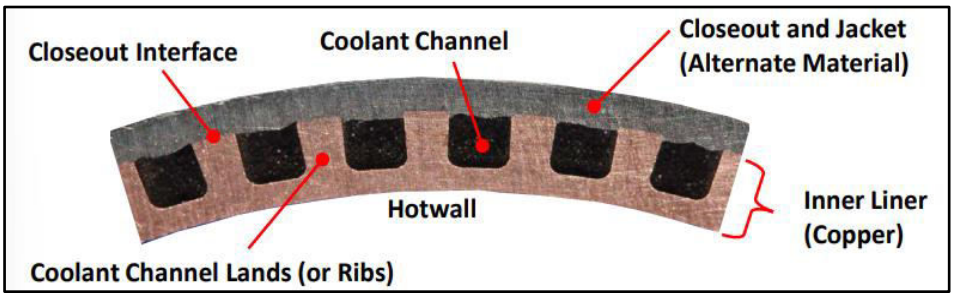


Figure 1: *Configuration of Bimetallic Channel Wall Nozzle*

NASA has been researching a range of advanced manufacturing methodologies for Channel Wall Nozzle (CWN) fabrication. A number one approach that has been appraised and demonstrated is the Laser Wire Direct Closeout (LWDC) technology. The figure shows a piece of a channel wall nozzle that comes with integral coolant channels, within an inner copper liner employing a bimetallic structure.
The channel wall configuration requires that the thickness of the hot wall be tightly controlled during the machining of the coolant channels. These channels are additionally closed out by bonding a closeout or structural jacket to the lands of the channels within the inner liner to contain the pressurized coolant within each channel. Inlet and outlet manifolds are fabricated individually and joined by a welding or brazing process to complete the nozzle. Channel wall nozzles offer cost and schedule savings thanks to minimal manufacturing steps and labour, as compared the tube- wall nozzles

varied accoutrements combinations are tested but the most effective is to possess monolithic accoutrements (which can be cast as single pieces & with no need for joints), which reduces the joining complexity. Bimetallic or multimodally channel wall nozzle-

1) Generally, have a copper liner
2) Material can vary radially or axially for-
    a) Weight optimization
    b) Increased thermal margins
    c) Increased structural margins

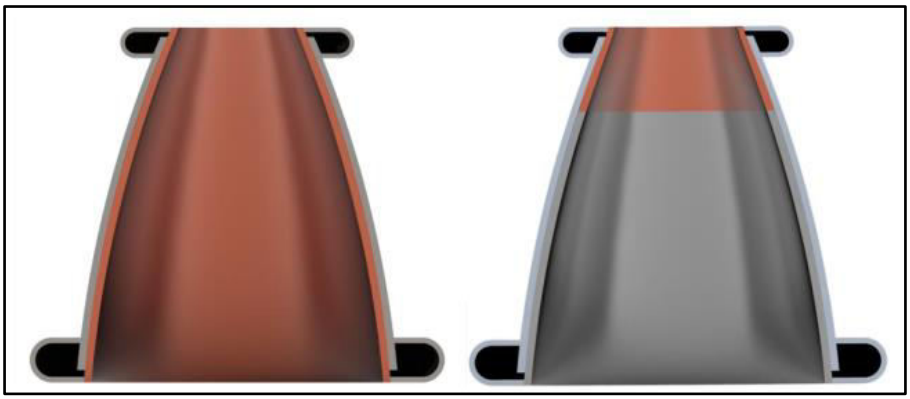


Figure 2: (Left) Radial bimetallic configuration and (Right) Axial/Radial Bimetallic/Multi-metallic CWN Configuration

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

The left image has a Radial Bimetallic configuration which uses copper- amalgamation for the complete length. For the closeout layer/ structural jacket, an alternate material (stainless- based or superalloy) was chosen.

- 2nd Material was applied in the radial direction for closeout.
- Uses copper as liner &non-copper material for channel closeout/ jacket to optimize weight.

The right image shows the AxialMulti-Alloy nozzle. As we go down axially, the warmth flux across the wall reduces. therefore, we will have copper amalgamation at the forward end of the nozzle (which has the highest heat flux) & also we will transition to a lower viscosity material which is nice enough for that low heat flux. Again, for the closeout/ jacket we will use an alternate material (stainless- based or superalloy).

- Uses varying accoutrements along the length of the hot wall to optimize for weight

a) Copper for prime heat flux- region
b) 2nd or 3rd material used for channel bonus (high strength jacket)

III. LWDC PROCESS

Laser Wire Direct Closeout (LWDC) is an additive manufacturing methodology that locally welds filler wire to the channel ribs & provides a structural jacket fabricated “in place”.
The process deposits a filler wire material that bridges the span of the coolant channels with no internal channel fillers necessitated. An independent wire feed and neutralize inert gas- purged beam melts the feedstock wire using a neighbourhood of fabric before the commencement of the coolant channels (generally progressing after end to the forward).
While the nozzle is rotated about the middle axis, the wire is deposited- piercing and joining to the preliminarily deposited layer (or area of stock) with a minor quantity of laser energy being employed to fuse the wire to the backside of the channel lands (or ribs). This process is repeated along the wall of the nozzle at continuously varying angles until the asked area is closed out. LWDC is employed for the direct closeout of the coolant channels and operation of the structural jacket.
Exemplifications of the LWDC process are frequently seen in Figure 3. The individual layers of closeout are observed covering axial along with the channel lands.

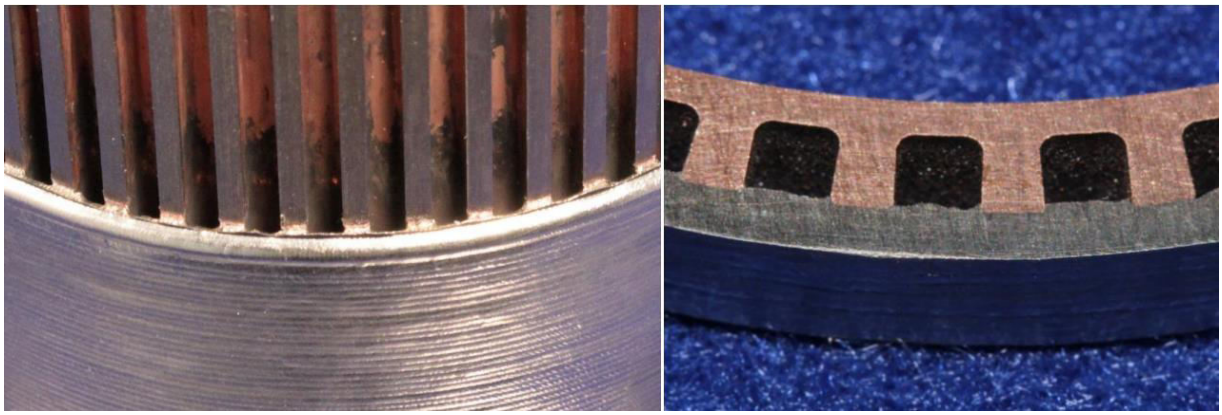


Figure 3: *Examples of Bimetallic LWDC Technology*

LWDC is additive manufacturing (AM) wire- fed laser sediment process that eliminates the necessity for a decent forbearance structural jacket and plating operations compared to traditional manufacturing. This process provides an immediate closeout of the coolant channels and forms the jacket integral within the same process. A low diameter wire is employed and also the low heat flux freeform wire- sediment process provides the power to make the jacket in situ while maintaining the geometry of the thin-walled channel lands or ribs, minimizing overall deformation.
The LWDC process is initiated by depositing material within the stock on the after end of the channel wall nozzle liner. Several passes are deposited during this region of stock before the axial stations where the coolant channels begin. This sediment within the starting region of stock provides for a starting" step” for posterior layers. For the closeout of the coolant channels, an off- axis ray of light and an off- axis wire- fed system is employed at angles defined relative to the nozzle wall half- angle and rotated a limited centres axis. A majority of the laser energy is targeted (spot size) on the preliminarily deposited" step “of fabric, while the remaining energy is targeted on the channel land. This enables the

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

fabric to enter the former layer while the material is likewise communed to the lands without burning through the lands or material dropping into the channels. An illustration of the LWDC process is seen in Figure. 4. The angles of the laser and wire- feed are continuously varied as a function of the nozzle exterior wall to forestall drop through and maintain the correct bonds. Overheating can invoke distortion of the liner wall or possible blow-through of the hot wall, so a mandrel is used.

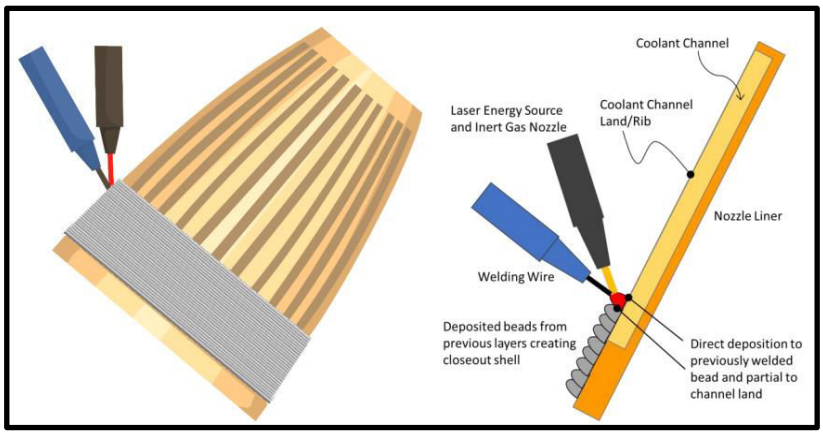


Figure 4: LWDC Process Overview

The primary advantage of the LWDC process is that the jacket and channel closeout are integrally formed, so forbearance are consequential looser compared to brazing or other laser-welded closeout processes. an endless bond is made at each of the ribs to make sure structural circumferences are met. excluding the necessity for channel fillers reduces post-processing time. The process does use small wire for sediment to regulate heat input into the part, and sediment rates are significant slower compared to other direct energy deposition (DED).
Still, now is neutralize by the elimination of a closeout jacket and posterior adherence operations.

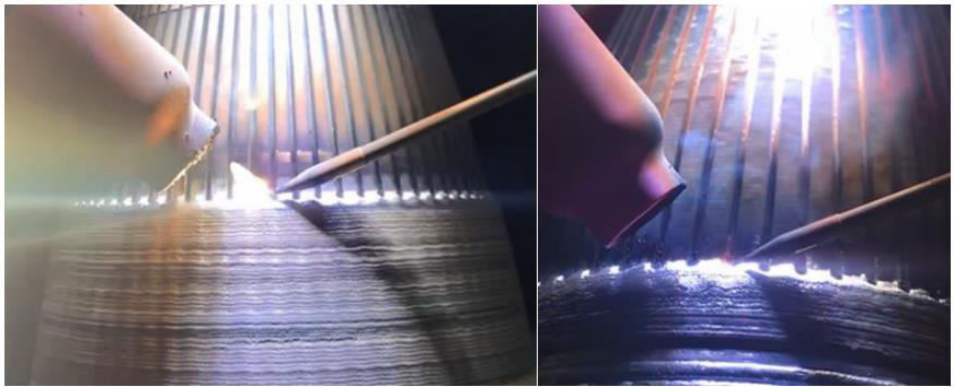


Figure 5: Nozzle Closeout being completed with LWDC

IV. CONFIGURATIONS TESTED

Nozzle 1 - 4: Monolithic configurations were tested previously.
The following series of subscale nozzles are fabricated for fire tests via LWDC bimetallic and multi-metallic processes.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

Unit	Configuration	Liner	LWDC Closeout	Heat Treatment
Nozzle #6	Radial Bimetallic, Intermediate Alloy	C-18150	Monel 400	Solution and Age
Nozzle #7	Axial Multi-alloy	C-18150 / CRES 347	Monel 400	Solution and Age
Nozzle #8	Radial Bimetallic, Direct	C-18150	Monel 400	Solution and Age
Nozzle #9	Radial Bimetallic, Direct	C-18150	Monel 400	None

Closeout/Structural jacket → Same for all nozzles (i.e., Monel 400)

Radial bimetallic → Have the same liner obviously (i.e., C-18150 ⇒ Cu-Cr-Zr)
Axial Multi alloy → This has 2 diff liners

C-18150: Forward end where heat flux is high
CRES 347: For the low heat flux-region

V. HOT FIRE TESTING AND RESULT

The testing was done on LOX & Gaseous Hydrogen (LOX/GH2) thrust chamber assembly. Thrust chamber assembly used a Powder Bed Fusion (PBF) additively manufactured coaxial injector and PBF additive manufactured GRCop-42 combustion chamber liner. The testing initially used water cooling to characterize the overall heat load of the thrust chamber assembly & eventually transitioned to full regenerative cooling using GH2. The injector included a middle port for the igniter.



Figure 6: Hot Fire Testing machine

A total of 72 tests were completed with a total test time of 3500 seconds.
Chamber pressure = 1225 psig/pound per square inch = 8.446078 MPa
Mixture Ratio = 8.0

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104108 |

Nozzle	Configuration	Coolant	Peak Chamber Pressure (psig)	Peak MR	Starts	Accumulated Time (sec)
Nozzle #6	Radial Bimetallic, Intermediate Alloy	GH2	1,122	6.2	3	540
Nozzle #7	Axial Multi-alloy	N/A	N/A	N/A	N/A	N/A
Nozzle #8	Radial Bimetallic, Direct	GH2	1,139	6.2 - 8.0	60	1,830
Nozzle #9	Radial Bimetallic, Direct	Water / GH2	1,225	6.2	9	1,130

Table: *Bimetallic LWDC nozzle fire test results.*

Nozzle	Test	Pc (psig)	MR	ΔP (psid)	ΔT (F)	Total Q (BTU/s)	R
Nozzle #6	PI100-013	1,122	6.13	86	111	295	129
Nozzle #8	PJ038-026A	1,118	6.14	142	112	295	218
Nozzle #9	PJ038-001	1,109	6.11	128	104	279	192

Table: *Performance Results*

While overall the nozzle hardware performed for sure, there have been several observations made about the test series relative to the various nozzle configurations. A summary of comparison tests with similar MR and Pc is shown within the above table. For every test, the performance of the nozzles was tabulated before setting the conditions for the following testing. Instrumentation included inlet and outlet temperatures and pressures within the manifolds yet as tubing. Additional backside (coldwall) thermocouples were tack welded to the nozzle jacket.

VI. CONCLUSION

In the future, there'll be various products around us that may be manufactured in space. For this, there'll be an enormous need for the launch vehicle's production to satisfy all the wants. Making profits, meeting deadlines for launch & ensuring the safety of both payload & crew members with advanced manufacturing methods, is extremely important for any launch service provider.

Designing a vehicle is extremely easy compared to manufacturing & thus, to simplify work & improve accuracy, additive manufacturing is getting used in various forms for prime standard products. Additive manufacturing is employed for product designing, process parameters monitoring, and relating the input parameters with the properties of the output which is the final part produced.

In 3D printing, it's very difficult to optimize operating process parameters, as they're not very linear in nature. This research, however, can efficiently be completed using an Additive manufacturing model, with relevance to its nonlinear nature. During this context, we came to grasp how 3d printing technology is often accustomed to making rocket nozzles.

REFERENCES:

1. Ражабов, Ш. Б. (2016). Использование современных информационных и педагогических технологий в деятельности тренеров-преподавателей. Актуальные проблемы физической культуры и спорта, 424-426.
2. Kabilov, A., Rajabov, S. B., Urmanov, B. N. (2021). Problem of Optimum Control Connected with Environmental Problems. The 5th International Conference on Future Networks & Distributed Systems, 733-737.
3. Очилов, С., Раджабов, Ш. Б., Омонов, А. А. (2021). Оптимизация времени прохождения нелинейной системы с параметром через область. Современные проблемы дифференциальных уравнений и смежных разделов, 1(1), 340-342.
4. Xashimxodjayev, S. I., Sadinov, A. Z., Rajabov, S. B. (2021). Methods of automation and management of waste recycling in the digital economy. Academic research in educational sciences, 2(CSPI conference 3), 149-154.
5. Urinovich, K. A., Qobiljonovich, R. O., Baxtiyorvich, R. S., Abdulakhatov, M. M. (2021). Modern content and concept of digital economy. ACADEMICIA An International Multidisciplinary Research Journal, *11*(11), 829-832.
6. Kobilov, A.U., Rikhsimboev, O.K., Rajabov, S. B. (2021). A global approach to assessing competitiveness digital economy. Экономика и бизнес теория и практика, 11-2, 115-119.
7. Кобилов, А. У., Тулаев, М. С., Ражабов, Ш. Б., Маматкодирова, Н. У. (2021). Правовая основа формирования цифровой экономики в республике узбекистан. Экономика и социум, 12(91), 96-104.
8. Yo'lchiboyeva, D. (2021). Maktabgacha ta'lim tashkilotlarida bola shaxsiga yo'naltirilgan ta'lim dasturini amaliyotga tatbiq etishning samarali usullari. XALQ TA'LIMI, 1(maxsus), 115-119.
9. Юлчибоева, Д. (2020). Ёш авлодга иқтисодий тарбия бериш ва тежамкорликни тарбиялаш – долзарб вазифа. ORIENTAL ART AND CULTURE, 1(1), 143-146.
10. Ҳайдаров, О., Юлчибоева, Д. (2020). Рефлексия психологик феномен сифатида. ORIENTAL ART AND CULTURE, 1(1), 120-123.
11. Yulchiboyeva, D. E. (2020). Bo'lajak o'qituvchilarning psixologo-pedagogik bilimlarini integratsiya jarayoni. Toshkent davlat pedagogika universiteti ilmiy axborotlari ilmiy-nazariy jurnali, 2(1), 134-138.
12. Юлчибоева, Д. Э. (2020). Эмоционал соҳа психологик ҳодиса сифатида. SCIENCE AND EDUCATION, 1(3), 701-707.
13. Alievich, E. E., Ergashevna, Y. D., Murodjonovich, K. A. (2020). Psychological features of the formation of perceptions of the world of professions in children of preschool age. European Journal of Research and Reflection in Educational Sciences Vol, 8(4), 61-63.
14. Юлчибоева, Д. Э., Ҳайдаров, О. М. (2019). Оила билан олиб бориладиган ижтимоий-педагогик фаолиятнинг айрим жиҳатлари. MODERN SCIENTIFIC CHALLENGES AND TRENDS, *2*(2), 71-76.
15. Khayrullayeva, R. L. (2021). Ways and functions of students speech development. ACADEMICIA An International Multidisciplinary Research Journal, 11(12), 389-393.
16. Azimova, F. P. (2019). Priority directions for development of the silk industry in the Republic of Uzbekistan. American Journal of Economics and Business Management, 2(4), 107-114.
17. Азимова, Ф. П., Қўзиев, С. А. (2022). Жаҳон мамлакатлари кластерлари ва Ўзбекистондаги кластерларнинг келгусидаги истиқболлари. Логистика ва иқтисодиёт илмий электрон журнал, 1(1), 87-94.
18. Azimova, F. P., Yuldashev, S., Akbarxujayeva, N. (2021). Analysis of the effectiveness of the textile industry tashkent institute of textile and light industry. Journal of Hunan University (Natural Sciences), 12(48), 1587-1597.
19. Azimova, F. P., Yuldashev, S., Akbarxujayeva, N. (2021). Analysis of the effectiveness of the textile industry. Vidyabharati International Interdisciplinary Research Journal, 2(13), 18-26.
20. Normuminovich, Y. S., Payziyevna, A. F., Azimdjanovna, A. N. (2021). Analysis of the effectiveness of the textile industry. 湖南大学学报 (自然科学版), 48(12), 1587-1597.
21. Payzievna, A. F. (2020). Improving Methodology to Evaluate the Efficiency of Textile Enterprises. Asian Journal of Technology & Management Research (AJTMR), 10(2), 23-26.
22. Азимова, Ф. П. (2020). Тўқимачилик корхоналарида баланслаштирилган кўрсаткичлар тизимини қўллаш орқали фаолият самарадорлигини бошқариш. БИЗНЕС-ЭКСПЕРТ ЖУРНАЛ ИЛМИЙ-ЭЛЕКТРОН ЖУРНАЛ, 1(12), 46-49.
23. Mutalova, D. M., Azimova, F. P., Yuldasheva, M. (2020). Some issues on perfection of the tax system of the Republic of Uzbekistan. International Journal of Psychosocial Rehabilitation, 24(2), 637-642.
24. Sultanov, T. M. (2021). Boshlang'ich sinflarda kombinatorika elementlarini qo'llash. Academic research in educational sciences, 2(9), 313-315.
25. Sultanov, T. M. (2022). Factors for the Development of Intellectual Ability of Future Teachers. International Journal of Innovative Research in Science Engineering and Technology, 11(3), 2346-2349.
26. Sultanov, T. (2022). Teaching Mathematics in Kindergarten and Primary Schools. International Journal Of Multidisciplinary Research In Science Engineering and Technology, 5(3), 376-377.
27. Sultanov, T. (2021). Boshlang'ich matematika kursi nazariyasi fanini o'qitishda Al-Xorazmiy ijodidan foydalanish. Mugalim ham uzluksiz bilimlendiru, 5(1), 23-26.
28. Sultanov, T. M. (2021). Boshlang'ich sinfda sinfdan tashqari ishlarni tashkil qilishda mental arifmetikasidan foydalanish metodikasi. Экономика и социум, 9, 967-968.
29. Хайруллаева, Р. (2022). АДИБЛАР ЗИКРИЛЛА НЕЪМАТ ИЖОДИ ХАҚИДА. Муҳаррир нашриёти, *1*(1), 1-64.
30. Xayrullayeva, R. L. (2021). Ona tili darsliklarida o'quvchilarning lug'at boyligini oshirishga samarali yondashilmoqda. Til ta'limining uzviyligini ta'minlashda innovatsion texnologiyalar muammo va yechimlar, *1*(1), 155-158.
31. Азимова, Ф. П. (2020). Тўқимачилик корхоналари фаолияти самарадорлигини бошқариш технологиясини такомиллаштириш. Бизнес-эксперт журнал илмий-электрон журнал, 1(7), 88-90.
32. Азимова, Ф. П. (2020). Тўқимачилик корхоналари фаолияти самарадорлигини баҳолаш. Иқтисодиёт ва таълим илмий журнал, 1(5), 29-34.

View publication stats

Annals of Hematology
https://doi.org/10.1007/s00277-022-04843-6

ORIGINAL ARTICLE

Mapping incidence and mortality of leukemia and its subtypes in 21 world regions in last three decades and projections to 2030

Rajesh Sharma[1] · **Chinmay Jani**[2,3]

Received: 13 November 2021 / Accepted: 5 April 2022
© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

Abstract

This study examines the burden of leukemia and its subtypes at the global, regional, and national levels in 21 world regions and 204 countries in the last three decades. The estimates of incidence, deaths, and age-standardized rates of leukemia for 21 regions and 204 countries for 1990–2019 were extracted from the Global Burden of Disease 2019 study. Average annual percentage change in 1990–2019 for 21 regions was utilized for projecting leukemia burden in 2030. Globally, there were 643,579 [586,980–699,729] incident cases and 334,592 [306,818–360,214] deaths in 2019 due to leukemia, up from 474,924 [388,559–560,550] cases and 263,263 [233,664–298,696] deaths in 1990. Between 1990 and 2019, the age-standardized incidence rate (ASIR) decreased from 9.6 [8.1–11.0] in 1990 to 8.2 [7.5–8.9] per 100,000 person-years in 2019, and the age-standardized mortality rate (ASMR) decreased from 5.8/100,000 [5.2–6.4] in 1990 to 4.3/100,000 [3.9–4.6] in 2019. Between 1990 and 2019, the ASIR decreased in majority of regions except Western Europe and high-income Asia Pacific, whereas the ASMR decreased in all 21 regions. In 2019, country-wise, the ASIR varied from 3.0/100,000 [2.3–3.7] in Palau to 35.1/100,000 [26.4–47.2] in San Marino and the ASMR spanned from 2.3/100,000 [1.7–2.8] in San Marino to 15.8/100,000 [12.0–20.4] in Syria. As per our projections, globally, there will be 720,168 incident cases and 367,804 deaths due to leukemia in 2030. Substantial improvements have been witnessed in leukemia mortality rates in all regions, especially high-income regions and countries. Health care policies focusing on diagnostic improvements, cancer registration, and newer therapeutics at reduced cost or with insurance coverage are needed in low and middle-income countries.

Keywords Leukemia · Global burden of leukemia · Incidence · Deaths

Introduction

Leukemia is a group of hematologic malignancies that can affect individuals in all age groups including children, adolescents, and adults. Leukemia burden has increased globally in terms of absolute counts but decreased in terms of age-standardized rates in the last three decades [1]. There are four main leukemia subtypes based on the speed of onset (acute versus chronic) and type of blood cells (myeloid versus lymphoid): acute lymphoid leukemia (ALL), chronic lymphoid leukemia (CLL), acute myeloid leukemia (AML), and chronic myeloid leukemia (CML). Out of these subtypes, ALL occurs primarily in children, AML occurs in both children and adults, whereas CLL and CML mainly occur in adults. In the last seven decades, leukemia diagnosis and treatment of a few subtypes have improved substantially, with 5-year survival rates of childhood ALL approaching 90% in high-income countries (HICs) [2]. At the same time, 5-year survival rates of CML has also improved substantially not only in HICs but also in low and middle-income countries (LMICs) [3].

To develop health care policies and strategies against leukemia, it is essential to gauge the current leukemia burden and its temporal patterns at the global, regional, and national levels. This study examined the changing trends of leukemia burden and its four major subtypes (AML, ALL, CML and CLL) at global, regional and national levels for 21 regions and 204 countries between 1990 and 2019. The estimates of leukemia and its four subtypes in terms of age, sex and

✉ Rajesh Sharma
rajesh.sharma@dtu.ac.in

[1] University School of Management and Entrepreneurship, Delhi Technological University, East Delhi Campus, Room No. 305, Vivek Vihar Phase II, Delhi 110095, India

[2] Mount Aubrun Hospital, Cambridge, MA, USA

[3] Harvard Medical School, Boston, MA, USA

Published online: 10 May 2022

location-wise incidence, mortality and their age-standardized rates (per 100,000 person-years) were extracted from the Global Burden of Disease 2019 study [4–6]. The temporal trends of absolute counts and age-standardised rates were examined at the global level and 21 regions. We also examined age-wise patterns of leukemia burden at the global level for 20 age groups and different leukemia subtypes in 2019. Lastly, we projected the leukemia burden to 2030 in 21 world regions. This comprehensive examination of temporal patterns of leukemia burden and its subtypes is expected to aid in policy formulation and help devise strategic interventions.

Materials and methods

The estimates of incidence and deaths at the global level for 21 regions and 204 countries for 1990–2019 were extracted from the Global Burden of Diseases, Injuries and Risk Factors 2019 Study (GBD 2019 study [4–6]) using online GBD results tool [7]. As per GBD 2019 study, leukemia burden is divided into four main categories (ALL, CML, AML, and CLL), with the uncategorized burden gets attributed to the category "other leukemia". The GBD estimates for leukemia pertains to 20 non-overlapping age groups: below-5, 5–9, 10–14, 15–19 …… 95-plus. The GBD estimation framework of incidence and mortality is detailed elsewhere [4–6]. Here, we briefly discuss the GBD methods employed to calculate the metrics examined in this study.

GBD estimation framework begins with the estimation of cancer-specific mortality [4]. Firstly, mortality-to-incidence (MI) estimates are generated using spatiotemporal Gaussian process regression (ST-GPR). These MI estimates are multiplied with incidence data from cancer registries to generate crude mortality estimates. The final mortality estimates are generated using the cause of death ensemble model (CODEm), in which crude mortality estimates from the last step are taken as inputs along with other covariates. The final incidence estimates are generated by dividing mortality estimates from the CODEm model with MI estimates from the ST-GPR model for each neoplasm, sex, location, and age group.

The age-standardized rates are produced using GBD standard reference population and are expressed in terms of counts per 100,000 person-years [4, 5]. All the estimates are provided with 95% uncertainty intervals (UIs); these UIS were derived from 2.5th and 97.5th percentile of 1000 draws [4–6] and reflect uncertainty arising from measurement error, systematic biases, data availability and modeling. These UIs are presented inside square brackets alongside mean estimates and in the places where we calculated or provided a single figure, that was based on the mean estimate of 1000 draws.

We calculate the lifetime risk (probability) of leukemia incidence and death using following formula.

$$\text{Cumulative Risk} = 1 - exp^{-CR}$$

CR denotes cumulative rate, which is calculated as $\sum_i Rate_i * t$, where $Rate_i$ denotes incidence or death rate in the ith age group, and t is number of years in the age group. As we considered, 5-year non-overlapping age groups, t is 5 for each age group.

To understand leukemia burden in different countries at different economic levels, we categorized all countries under four categories as per World Bank Income Classification: low-income (LI), lower-middle income (LMI), upper-middle income (UMI), and high income (HI). The thresholds of gross national product per capita in 2020 for categorizing countries are as follows: LI(< $1045), LMI($1046–$4095), UMI($4096–$12695) and HI(> $12696) [8].

Statistical analysis

Total number of leukemia cases and deaths in 2020–2030 was projected using average annual percentage change (AAPC) in the last three decades (1990–2019). The AAPC for various metrics was calculated based on the following regression.

$$\ln(Y) = \alpha_0 + \alpha_1 T$$

Y is the metric for which AAPC is calculated in the above equation and T is time (years). The AAPC is calculated as per the following formula.

$$\text{AAPC} = \exp(\alpha_1) - 1.$$

The burden of leukemia (number of cases and deaths) was projected into the future (2020–2030) using the AAPC calculated in the last step as per the following formula.

$$Y_{t+i} = (Y_t)^{AAPC+i}$$

In the above formula, i varies from 1 (2020) to 11 (2030). In this study, we compared the incident cases and deaths in 2030 with 2019 for 21 world regions.

Results

Temporal patterns of leukemia at global level

Globally, there were 643,579 [586,980–699,729] incident cases and 334,592 [306,818–360,214] deaths due leukemia in 2019, increasing from 474,924 [388,559–560,550] incident cases and 263,263 [233,664–298,696] deaths in 1990 (Fig. 1). Between 1990 and 2019, age-standardized incidence rate (ASIR) of leukemia decreased from 9.6/100,000

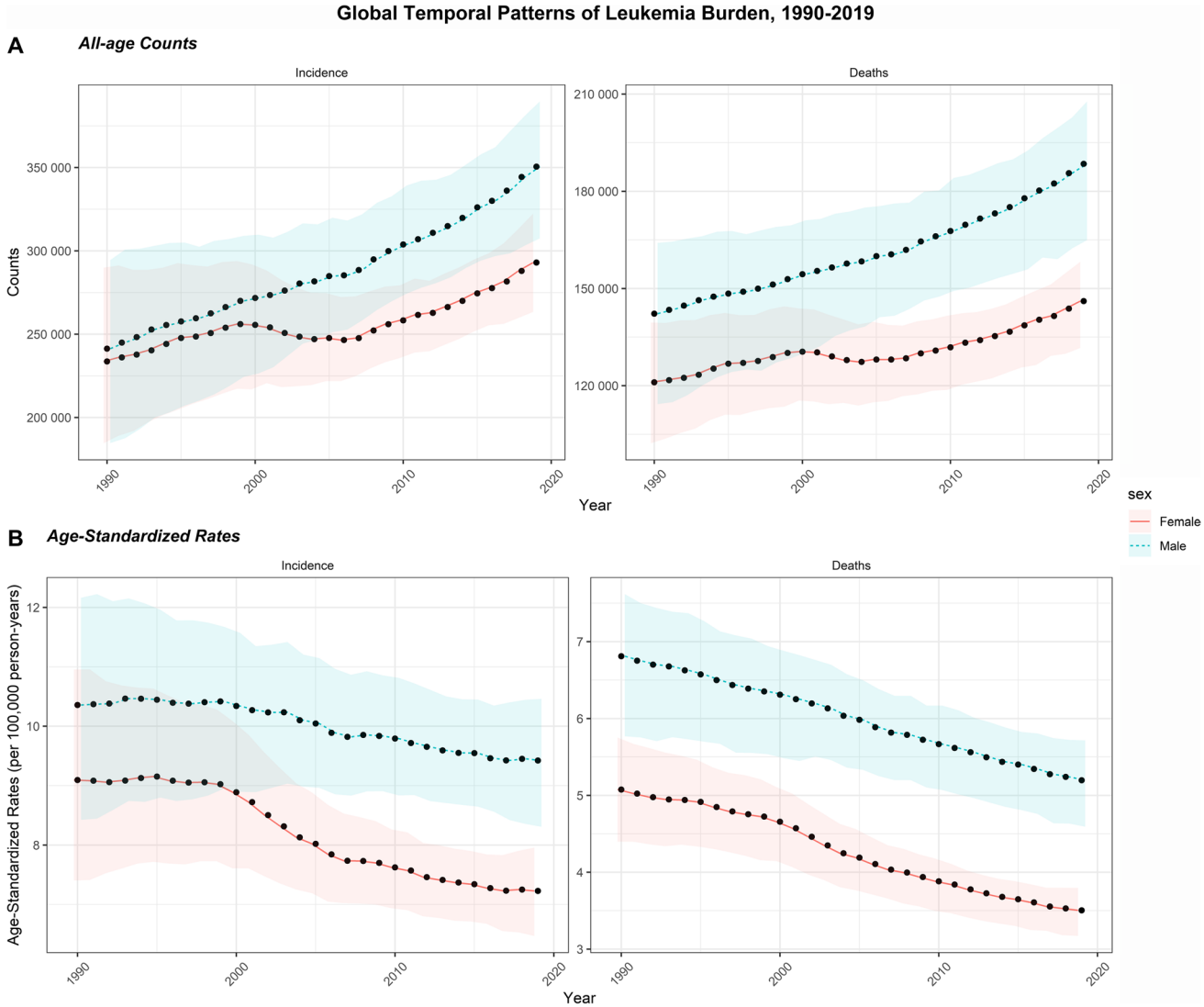


Fig. 1 Global temporal patterns of leukemia burden, 1990–2019. **a**) All-age counts. **b**) Age-standardized rates. Incidence: Incident Cases; Deaths: Death Count. The age-standardized rates of incidence and death were expressed as counts per 100,000 person-years. Data Source: Global Burden of Disease 2019 Study

[8.1–11.0] in 1990 to 8.2/100,000 [7.5–8.9] in 2019, and age-standardized mortality rate (ASMR) decreased from 5.8/100,000 [5.2–6.4] in 1990 to 4.3/100,000 [3.9–4.6] in 2019 (Fig. 1).

Supplementary Figure S1 presents temporal patterns of age-standardized rates of leukemia subtypes at the global level. Between 1990 and 2019, the ASIR (per 100,000 person-years) of ALL increased from 1.2 [1.0–1.6] to 2.0 [1.6–2.2], CLL increased from 1.1 [1.0–1.1] to 1.3 [1.2–1.5], whereas ASIR of CML decreased from 1.0 [0.8–1.2] to 0.8 [0.7–0.9] and other leukemia decreased from 4.9 [3.4–6.0] to 2.6 [2.2–2.9]. Globally, against the rise in ASIR, the ASMR of AML decreased between 1990 and 2019 (Supplementary Figure S1). Both ASIR and ASMR of CML increased up to 2000 and decreased after that.

Sex-wise burden of leukemia at global level in 2019

Males accounted for 54.5% (350,582 [307,570–389,660]) of total leukemia cases and 56.3% (188,469 [164,929–207,722]) of leukemia deaths in 2019. For both sexes combined, ASIR of leukemia was 8.2/100,000 [7.5–8.9], ranging from male ASIR of 9.4/100,000 [8.3–10.5] to female ASIR of 7.2/100,000 [6.5–8.0] (Supplementary Table S1). The ASMR for both sexes combined stood at 4.3/100,000 [3.9–4.6], varying between 3.5/100,000 [3.2–3.8] in females to 5.2/100,000 [4.6–5.7] in males with an incidence rate ratio (IRR) of 1.3 and mortality rate ratio (MRR) of 1.5.

For leukemia subtypes, male–female disparities were similar, with males accounting for more cases and deaths for all leukemia subtypes except CML(Supplementary Table S1). In case of CML, cases in males (55,283 [48,869–66,521]) were greater than females (48,184 [42,787–55,237]), however, CML deaths were estimated to be equal between males (22,306 [19,988–26,345]) and females (22,306 [19,858–25,084]). Male age-standardized rates were also greater than female age-standardized rates with IRR varying between 1.3 and 1.4 and MRR varying between 1.4 and 1.7 for different leukemia subtypes.

Table 1 Region wise burden of leukemia in 2019

Region	Incidence	Deaths	ASIR	ASMR
Andean Latin America	4,423 [3,234–5,622]	3,045 [2,259–3,807]	7.3 [5.3–9.3]	5.2 [3.8–6.4]
Australasia	4,717 [3,814–5,799]	2,367 [2,121–2,571]	10.5 [8.4–12.8]	4.8 [4.4–5.2]
Caribbean	3,607 [2,835–4,446]	2,364 [1,955–2,814]	7.6 [5.8–9.5]	4.8 [3.9–5.8]
Central Asia	5,247 [4,606–6,048]	3,188 [2,849–3,604]	6.0 [5.3–6.9]	3.9 [3.5–4.4]
Central Europe	17,318 [15,237–19,601]	9,586 [8,459–10,786]	9.6 [8.4–10.9]	4.9 [4.3–5.4]
Central Latin America	15,977 [13,603–18,617]	10,721 [9,283–12,353]	6.6 [5.6–7.7]	4.4 [3.9–5.1]
Central Sub-Saharan Africa	4,091 [3,061–5,533]	2,417 [1,841–3,061]	3.9 [2.9–5.0]	3.0 [2.2–4.0]
East Asia	159,358 [131,904–185,877]	62,983 [52,697–73,469]	10.4 [8.7–12.3]	3.7 [3.1–4.3]
Eastern Europe	21,615 [19,528–23,850]	11,746 [10,640–12,897]	7.5 [6.9–8.3]	3.9 [3.5–4.3]
Eastern Sub-Saharan Africa	21,548 [13,995–29,460]	12,064 [8,398–15,924]	5.9 [4.2–7.7]	4.4 [3.2–5.6]
High-income Asia Pacific	29,196 [24,933–33,271]	12,117 [10,459–13,028]	10.3 [9.0–11.7]	3 [2.7–3.2]
High-income North America	61,549 [53,738–70,238]	34,706 [32,027–36,512]	10.7 [9.4–12.1]	5.7 [5.3–5.9]
North Africa and Middle East	39,297 [32,617–45,056]	25,143 [21,109–28,826]	7.8 [6.5–8.8]	5.4 [4.6–6.1]
Oceania	808 [533–1,225]	457 [330–644]	6.8 [4.8–9.7]	4.8 [3.6–6.5]
South Asia	59,856 [51,907–70,105]	44,551 [38,745–52,631]	3.8 [3.3–4.5]	3.0 [2.6–3.5]
Southeast Asia	42,263 [35,843–49,692]	28,502 [24,415–32,956]	6.8 [5.8–8]	4.7 [4.1–5.5]
Southern Latin America	5,387 [4,273–6,700]	3,840 [3,589–4,063]	7.2 [5.7–8.9]	4.9 [4.6–5.2]
Southern Sub-Saharan Africa	2,763 [2,383–3,135]	1,956 [1,651–2,174]	4.3 [3.7–4.9]	3.4 [2.8–3.7]
Tropical Latin America	12,371 [11,557–13,082]	9,066 [8,445–9,504]	5.5 [5.1–5.9]	4.0 [3.7–4.2]
Western Europe	118,624 [102,901–135,441]	44,997 [39,920–47,622]	16.9 [14.7–19.4]	4.9 [4.5–5.1]
Western Sub-Saharan Africa	13,562 [10,265–17,281]	8,776 [6,932–10,831]	3.9 [3.1–4.7]	3.1 [2.5–3.7]

Incidence: All-age cases; Deaths: All-age deaths; ASIR: Age-standardized incidence rate (per 100,000 person-years); ASMR: Age-standardized mortality rate (per 100,000 person-years). 95% uncertainty intervals are presented inside square brackets. Data Source: Global Burden of Disease 2019 Study

Region-wise leukemia burden in 2019

Table 1 presents the leukemia burden (absolute counts and age-standardized rates) in 21 regions in 2019. In terms of absolute counts, East Asia was the leading region with 159,358 [131,904–185,877] incident cases and 62,983 [52,697–73,469] deaths in 2019. Western Europe (118,624 [102,901–135,441]) and high-income North America (61,549 [53,738–70,238]) were the second and third leading regions in terms of incident cases. In terms of leukemia deaths, South Asia (44,551 [38,745–52,631]) was ranked third behind East Asia (62,983 [52,697–73,469]) and Western Europe (44,997 [39,920–47,622]), ahead of high-income North America (34,706 [32,027–36,512]).

ASIR of leukemia was the highest in high-income regions led by Western Europe (16.9/100,000 [14.7–19.4]), followed by high-income North America (10.7/100,000 [9.4–12.1]) and Australasia (10.5/100,000 [8.4–12.8]) and ASMR was the highest in high-income North America (5.7/100,000 [5.3–5.9]), with mixed patterns in regions as per their development status (Table 1). For instance, low and medium-income regions such as North Africa and Middle East (5.4/100,000 [4.6–6.1]) and Andean Latin America (5.2/100,000 [3.8–6.4]) had higher ASMR than Western Europe (4.9/100,000 [4.5–5.1]), although their ASIRs were much lower than Western Europe (Table 1).

Figure 2 presents region-wise ASIR and ASMR for 21 regions for different leukemia subtypes in 2019. The ASIR of ALL was the highest in high-income regions (e.g., Western Europe 6.3/100,000; high-income Asia Pacific 5.4/100,000), whereas its ASMR was the highest in Central Latin America (1.5/100,000), followed by Andean Latin America (1.3/100,000). Although high-income regions (e.g., High-income North America, Central Europe) had the highest ASIR and ASMR of CLL, the gap between ASIR and ASMR was wider in these regions as compared to regions in Asia and Africa. In the case of CML, there was apparent disparity between regions in terms of gap between ASIR and ASMR. For instance, Western Europe had the highest ASIR (3.0/100,000), which is more than double as that of eastern sub-Saharan Africa (SSA) (1.2/100,000); however, the ASMR of Westrn Europe (0.3/100,000) was lower than eastern SSA (1.1/100,000). In almost all regions except high-income regions (e.g., Western Europe, high-income Asia Pacific, high-income North America), the age-standardized rates of "other leukemia" were greater than four main leukemia types (Fig. 2).

A few interesting observations emerge as per the absolute count of cases and deaths of leukemia subtypes in 21 regions in 2019 (Supplementary Figure S2). First, a substantial portion of leukemia burden was cornered by other leukemia (even more than all other types combined in a few regions) in regions of Asia (e.g., East Asia, Southeast Asia) and SSA (e.g., Eastern SSA, Western SSA) and North Africa and Middle East. Second, the percentage of burden of AML in total leukemia cases was lower, but its proportion in total leukemia

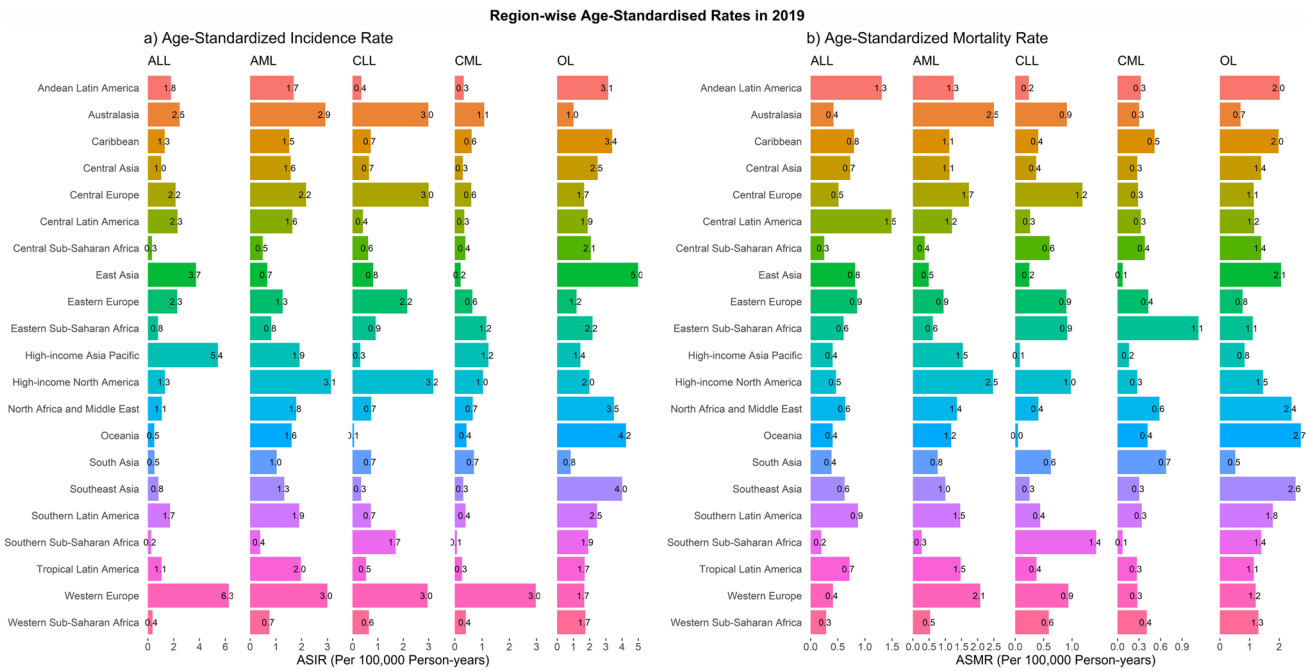


Fig. 2 Region wise age-standardized rates of four main leukemia types in 2019. **a**) Age-standardized Incidence Rate. **b**) Age-standardized Mortality Rate. The age-standardized rates are expressed in terms of counts per 100,000 person-years. ALL: Acute lymphoid leukemia; CLL: Chronic lymphoid leukemia; AML: Acute myeloid leukemia; CML: Chronic myeloid leukemia; OL: Other Leukemia. Data Source: Global Burden of Disease 2019 Study

deaths was higher in almost all regions. In contrast, CML and ALL accounted for a greater proportion of incident cases and a lower proportion of deaths, particularly in high-income regions (e.g., Western Europe, high-income Latin America, and high-income Asia Pacific).

Regionwise temporal patterns of leukemia burden, 1990–2019

Supplementary Figure S3 presents the temporal patterns of age-standardized rates of leukemia in 21 world regions. Between 1990 and 2019, ASIR of leukemia decreased in most regions except Western Europe and high-income Asia Pacific, whereas ASMR of leukemia decreased or stagnated in all 21 regions. In Western Europe and Australasia, ASIR remained almost stagnant in the last decade. In most regions, reductions in ASMR were faster than reductions in ASIR such that the gap between ASIR and ASMR of leukemia widenened over the years.

Supplementary Figure S4 presents temporal patterns of lifetime risk of leukemia incidence and death in 21 regions between 1990 and 2019. In regions with higher incidence rate in 1990 (e.g., Australasia, High-income North America, Western Europe, Central Europe), cumulative life time risk of leukemia incidence increased, whereas lifetime risk of leukemia death decreased between 1990 and 2019. In majority of regions, cumulative risk of leukemia incidence increased faster than lifetime risk of leukemia death with a few exceptions. For instance, the cumulative risk of leukemia incidence and death increased in consonance in Andean Latin America, Tropical Latin America and South Asia. In majority of SSA regions, lifetime risk of leukemia incidence and death changed parallelly, whereas in East Asia and Central Latin America, the lifetime cumulative risk of leukemia incidence grew faster than lifetime risk of death (declined or stagnated) in recent years.

Supplementary Table S2 compares the burden of leukemia (incident cases and deaths) in 2030 and 2019. As per our projections, there will be 720,168 incident cases and 367,804 deaths globally due to leukemia in 2030 and East Asia is projected to remain the leading region with 156,748 cases and 59,538 deaths in 2030. Central Asia and East Asia are the only regions expected to experience a reduction in leukemia incident cases between 2019 and 2030. Apart from Central and East Asia, Eastern Europe is the only other region projected to reduce leukemia deaths in 2030 compared to 2019.

Country-wise leukemia burden in 2019

In terms of incident cases, China (154,648 [181,075–127,207]), USA (53,107 [54,610–61,437]), and India (43,570 [37,235–52,165]) were the leading countries in 2019. In terms of death counts, China (60,381 [50,213–70,988]) was the leading country, followed by India (33,827 [28,764–40,416]) and USA (31,378 [28,944–33,015]) (Supplementary Table S3). There were wide variations in terms of age-standardized rates of leukemia in different countries (Fig. 3). ASIR was the lowest in Palau (3.0/100,000 [2.3–3.7]) and Morocco (3.1/100,000 [2.4–3.9]) and highest in San Marino (35.1/100,000 [26.4–47.2]). The ASMR was the highest in Syria (15.8/100,000 [12.0–20.4]), followed by Afghanistan (10.0/100,000 [7.0–14.4]) and lowest in Palau (2.3/100,000 [1.7–2.8]). Middle-to-high ASIRs were geographically concentrated in HICs such as the USA and parts of Europe, whereas middle-to-high ASMRs were scattered in parts of Asia, America, and Africa (Fig. 3).

In general, the age-standardized rates increased with income levels of countries with the HICs recording higher age-standardized rates than other income groups in 2019 (Fig. 4). The positive gradient between income and age-standardized rate was more apparent for ASIR than ASMR such that median ASIR of HICs was almost double than LICs, whereas there was lesser difference between ASMR of HICs and LICs. There are a few exceptions in different income groups; for instance, Afghanistan in LICs and Lebanon in UMICs had much higher ASIR than not only their respective income group, but their age-standardized rates were also much higher than higher income groups.

Age-wise burden of leukemia in 2019

Figure 5 illustrates the age distribution of leukemia in 20 non-overlapping age groups, and Supplementary Figure S5 shows the age distribution of four main leukemia types at the global level in 2019. For aggregate leukemia, the age distribution of absolute counts was U-shaped, with one peak in the below-5 age group and second peak in 70–74 age group. In contrast, age-specific rates continued to rise, and as a result, the oldest age groups had the highest age-specific rates (Fig. 4). There were contrasting patterns in the age distribution of leukemia subtypes, with ALL recording the highest number of cases and deaths in the youngest age group (Supplementary Figure S5). In contrast, there were no estimated cases in those below the age of 25 years in CLL. CML had one peak in the under-5 age group, but its substantial burden occured in adults and older age groups such that age-specific rates were the highest in oldest age groups. AML had a J-shaped pattern of incident cases and deaths, with a peak in the under-5 age group and another peak in the 70–74 age group and the highest age-specific rates in the oldest age groups.

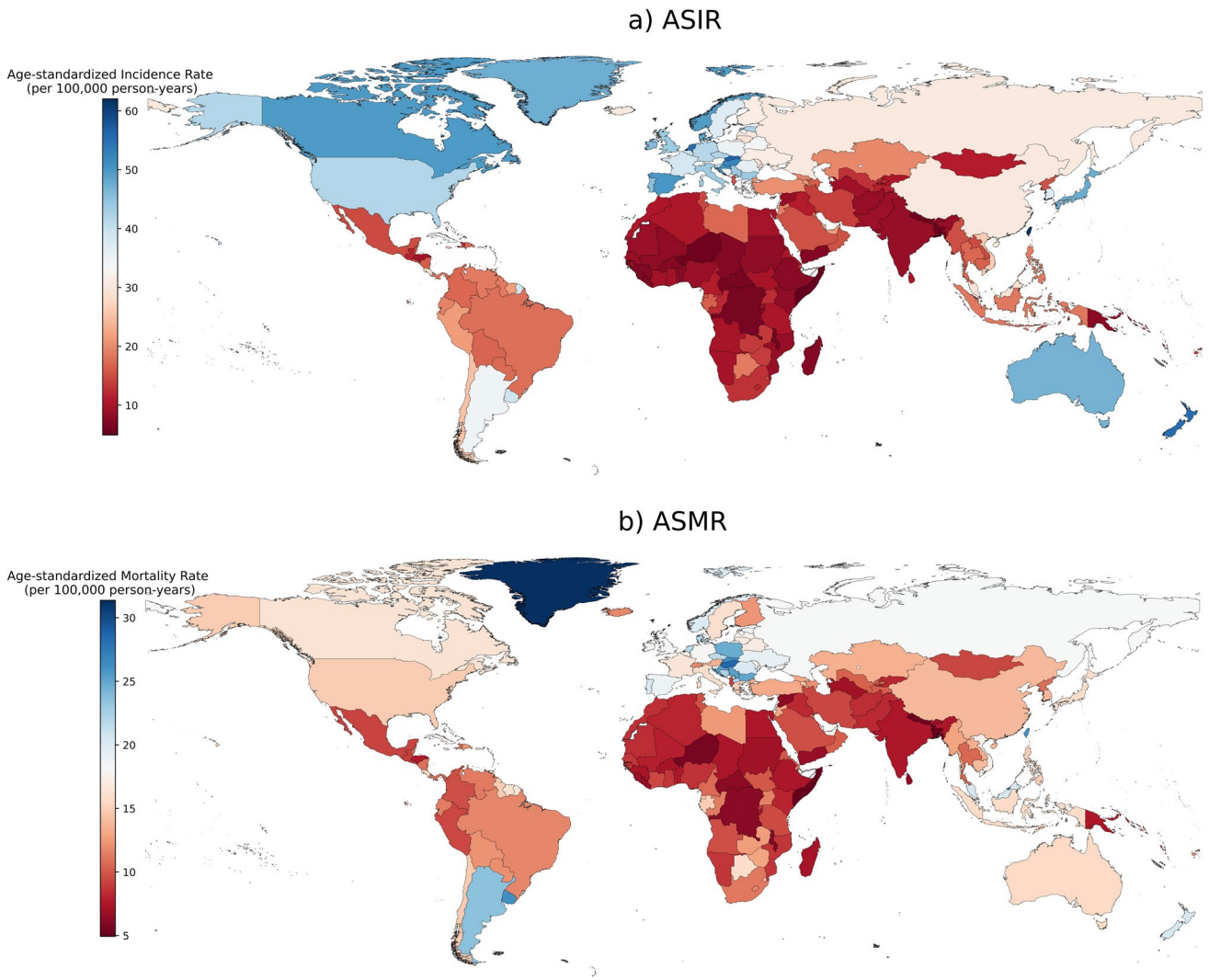


Fig. 3 Geographical patterns of age-standardized rates of leukemia in 2019. **a**) ASIR. **b**) ASMR. ASIR: Age-standardized incidence rate (per 100,000 person-years); ASMR: Age-standardized mortality rate (per 100,000 person-years). Data Source: Global Burden of Disease 2019 Study

Discussion

Between 1990 and 2019, incident cases and deaths due to leukemia increased globally; however, the age-standardized rates witnessed a downfall for leukemia and its subtypes. In 2019, Asian regions accounted for highest number of cases and deaths due to leukemia while the ASIR was the highest in high-income regions such as high-income North America, high-income Asia Pacific, and Western Europe. Notably, the ASMR of leukemia decreased in all 21 GBD regions, whereas ASIR increased in a few high-income regions. The decreasing trends in mortality rates can be attributed to substantial improvements in the advanced diagnostic and therapeutic options for leukemia subtypes such as ALL and CML, particularly in HICs [9].

Males accounted for a greater proportion of cases and deaths due to leukemia as well as its subtypes. The male age-standardized rates were 1.2–1.7 times that of females for different leukemia subtypes and higher burden among males than females have also been found in previous studies [10, 11, 12, 13]. Higher leukemia incidence in males might be linked to higher exposure to risk factors such as smoking and occupational exposure to carcinogens [14].

Tobacco smoking is classified as carcinogenic towards leukemia with sufficient evidence by International Agency for Research on Cancer (IARC) [15]. Even parental smoking is considered carcinogenic to children, although with limited evidence [15]. Previous research has also identified high body-mass index as one of the dominant risk factors of AML [16]. Moreover, childhood obesity is not only associated with risk of cancer during childhood but it may also lead to a

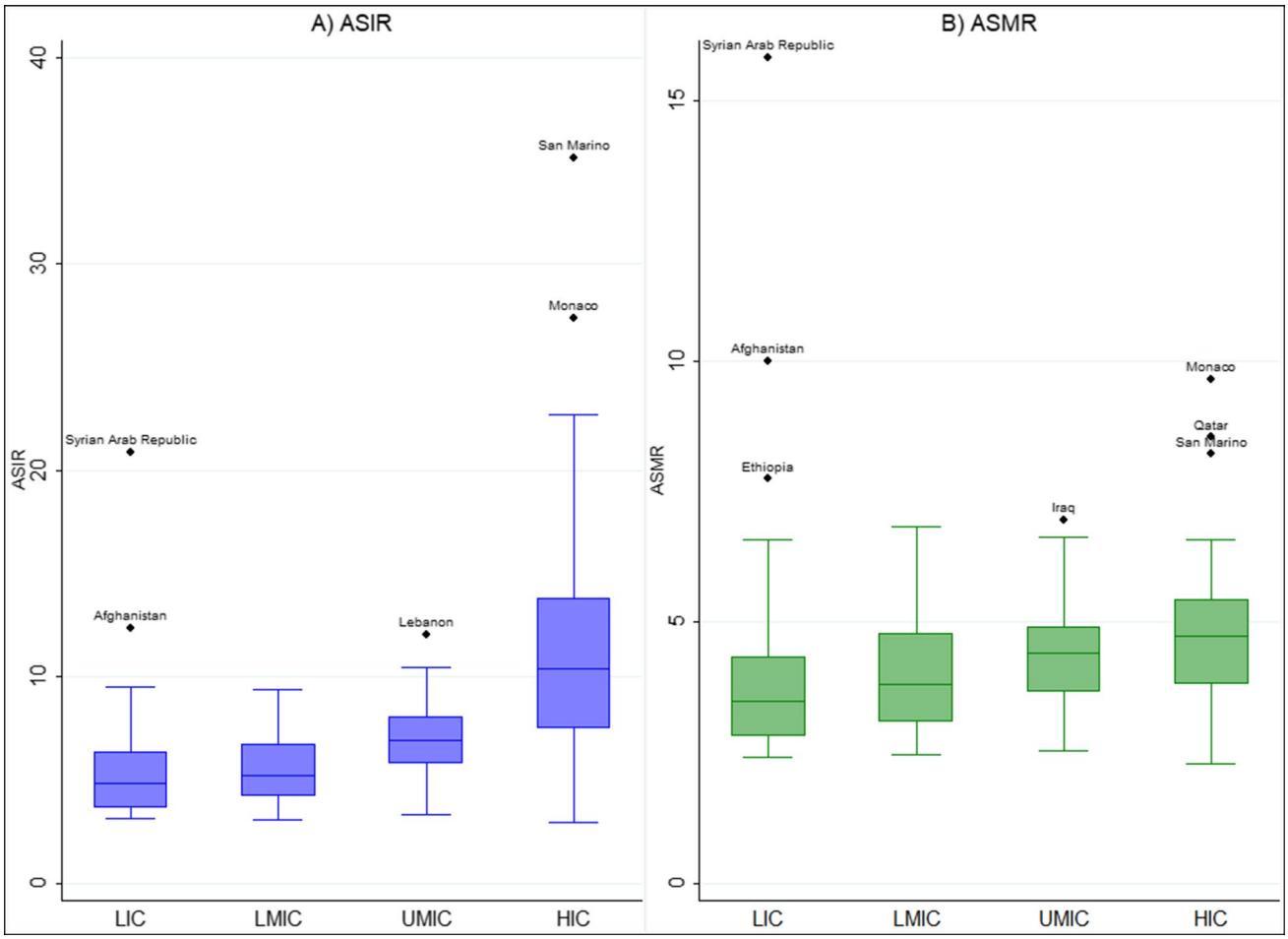


Fig. 4 Age-standardised rates of leukemia as per World Bank income status of countries in 2019. The countries were categorized in four categories as per World Bank Income Classification: low-income (LI), lower-middle income (LMI), upper-middle income (UMI) and high income (HI). The thresholds of gross national product (GNI) per capita for country classifications are: LI (< 1046), LMI (1046–4095), UMI(4096–12695) and HI(> 12695)

higher risk of cancers, including leukemia, in adulthood [17, 18]. Although there are a few avenues of leukemia prevention in children, a few studies have proposed deficiency of early-life microbial exposure, secondary mutations as a result of a dysregulated immune response or chronic inflammation to common infections, exposure to ionizing radiations in-utero or treatment of some other cancer and parental exposure to smoking as a few risk factors for leukemia in children [19, 20].

Globally, ASIR and ASMR of CML have decreased since early-2000s; however, there are disparities between regions at different levels of development. For instance, Western Europe had the highest ASIR (3.0/100,000), more than double as that of eastern SSA (1.2/100,000); however, ASMR of CML in Western Europe (0.3/100,000) was lower than eastern SSA (1.1/100,000). In HICs, Imatinib, a tyrosine kinase inhibitor (TKI), has significantly boosted the 5-year survival rates of CML to close to 90% [21]. Moreover, life expectancy of Imatinib-treated CML patients approaches that of the general population [22]. In LMICs, however, there are still challenges relating to delayed diagnosis, BCR/ABL testing, and availability and cost of second and third-generation TKIs [23]. In LMICs, tools such as Fluorescence in situ hybridization (FISH) and reverse transcription-polymerase chain reaction (RT-PCR), which can detect BCR/ABL aberration are also insufficiently available. High out-of-pocket expenditures and lack of insurance coverage (or Universal Health Coverage) also hinders cancer treatment in LMICs in Asia, Africa and Latin America, availability of Glivec (Imatinib) at low or no cost through programs such as Glivec International Patient Assistance Program (GIPAP) can significantly boost survival rates of leukemia [23]. In fact, a previous research study has found that GIPAP patients (receiving Imatinib free-of-cost in LMICs) have almost similar 5-year survival rates as HICs [24].

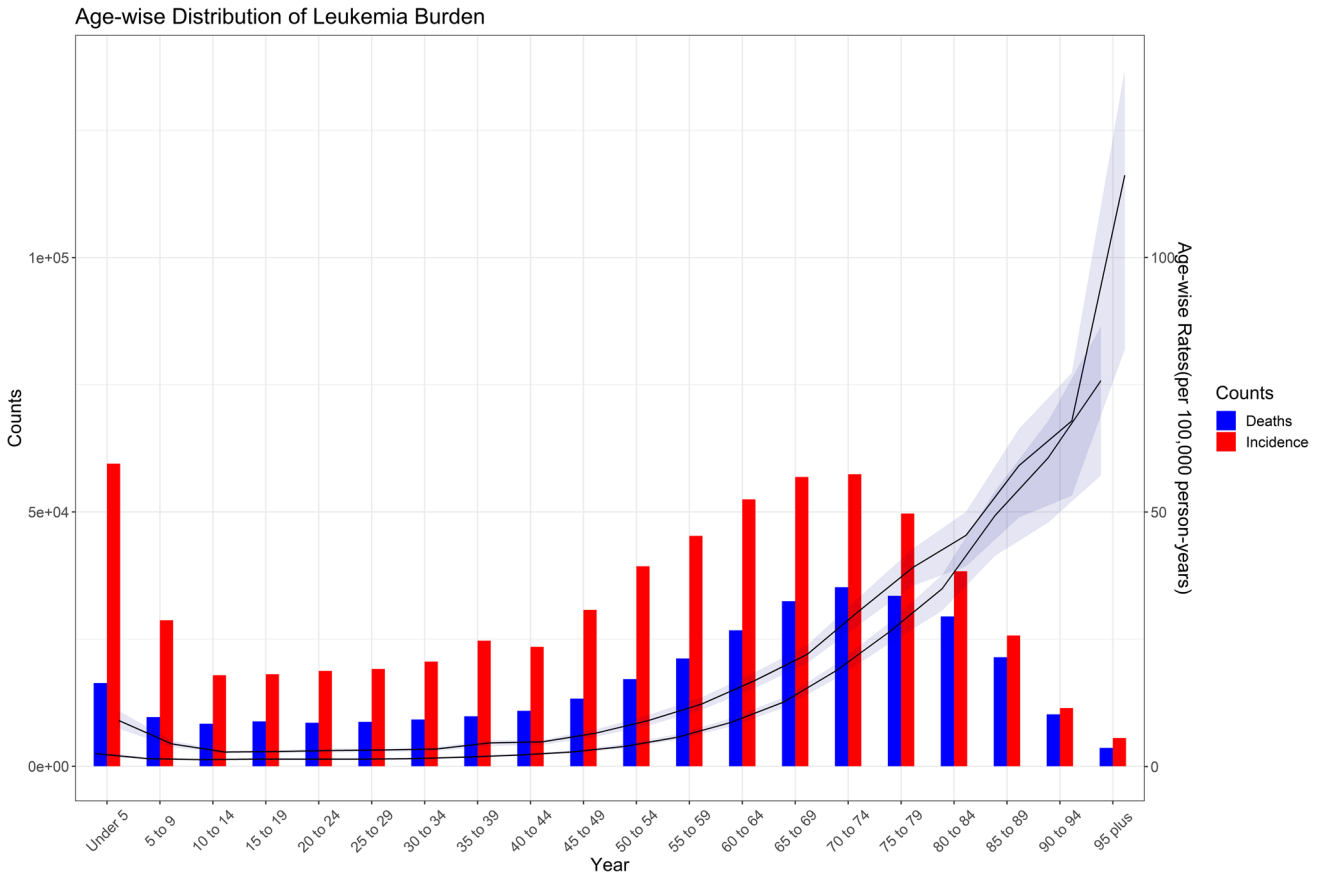


Fig. 5 Age distribution of leukemia burden in 2019. Deaths: Age-specific deaths; Incidence: Age-specific cases. The vertical bars represent absolute counts and line with grey band represents the age-specific rate per 100,000. Data Source: Global Burden of Disease 2019 Study

Similar to the reductions in ASMR of CML, the ASMR of ALL has also decreased substantially in majority of high-income regions. Since the chemotherapeutic regents initiated by Sydney Farber, leukemia medication has come a long way, with 5-year survival rates for childhood ALL reaching as high as 90% in HICs [2]. However, there are differences in 5-year survival rates in countries at different income levels. For instance, as per CONCORD-3, the 5-year survival rates of childhood ALL varied from 49.8% in Ecuador to 95.2% in Finland [25]. In LMICs, deaths due to drug toxicity, treatment abandonment and relapses remain significant challenges towards tackling childhood ALL [26–28]. Previous research has also identified that the children with ALL are 10-times more likely to acquire and die of infections in LMICs than HICs [29].

In the case of ALL, there are also differences in 5-year survival rates in different age groups. For instance, childhood ALL have better survival rates (close to 90%) than older age groups, with the 5-year survival rates decreasing with age, reaching approximately 25% among those above the age of 50 years [30, 31]. Better survival rates in childhood ALL might be due to a patient's biological characteristics, physiology, and immune system [32]. For instance, the Philadelphia chromosome (Ph+), primarily absent in childhood ALLs, is present in 20–30% of adult ALLs [33]. Notably, TKIs have significantly boosted disease-free survival and overall survival of Ph+ adult ALLs [34, 35].

AML accounted for a large percentage of leukemia cases and even a larger percentage of leukemia deaths in most regions. Even though AML treatments have substantially boosted outcomes for younger patients, the prognosis of AML remains poor for older patients [36], and 70% of AML patients older than 65 years are expected to die of AML within one year of diagnosis [37]. Risk factors associated with AML are prolonged exposure to petrochemicals, Benzene, pesticides, and ionizing radiations [38]. In addition to these risk factors, there is a risk of therapy-induced myelodysplastic syndrome and AML in childhood survivors of hematological malignancies or during the treatment of other solid malignancies [39].

The classification of AML requires multidisciplinary diagnostic tests, including immunohistochemistry, cytochemistry, or both, in addition to molecular genetics analysis [39]. Cytogenetic analysis, which looks for changes in chromosomes such as inversion, translocation, addition, or deletion, is crucial in the case of AML as a diagnostic tool and also has a prognostic value. For instance, a complex karyotype (≥ 3 abnormalities) is detected in 10–12% of AML patients has a poor prognosis [40, 41]. In LMICs such as Brazil, there is limited or no access to immunophenotyping, cytogenetic and molecular tests, drugs, beds for intensive chemotherapy along with delayed or no access to transplantation [42]. Even for childhood ALL, flow cytometry and cytogenetics can help identify high-risk disease such as T cell or hypodiploid ALL [43]; however, many of these diagnostic tools, which also have prognostic value, are scarce in several LMICs [44].

Leukemia survival rates can be boosted in LMICs through early diagnosis, timely and cost-effective treatments. In countries like India and those in SSA, particularly in the rural areas of these countries, leukemia patients or their parents often approach local or traditional healers [45], and flu-like symptoms sometimes further delay the diagnosis. Moreover, even if patients reach the diagnostic center, there may not be facilities such as Immunohistopathology and Immunophenotyping to ascertain and characterize the cancer appropriately [46]. Apart from these hindrances, high out-of-pocket expenditures, lack of universal health coverage, and distance from oncologic centers are major limitations to cancer care in LMICs [47]. Ramping up universal health coverage in LMICs can reduce therapy abandonment rates in these countries [45]. In Mexico, for instance, establishment of the Fund for Protection against Catastrophic Expenditures, which included ALL as one of the first interventions, has reduced the treatment failures, especially among poor and vulnerable sections of society [48, 49]. Hub and spoke models can also be adopted with more diagnostic facilities in rural areas and specialized treatment in select centers [50, 51]. As for collaborative arrangements, twinning arrangements between institutions in HICs and LMICs have shown success in improving the treatment of childhood malignancies, including leukemia [50–52].

Limitations

GBD estimates for leukemia are expected to be downward biased for several leukemia types, and the actual estimates are expected to be higher in LMICs of Africa, America and Asia. The reasons for this bias are under-reporting, under-ascertainment and misdiagnosis of leukemia cases and deaths in these countries [46, 53]. In the context of childhood cancers, Lam et al. [46] reported a substantial incidence deficit (actual-recorded) in LMICs, particularly in cancers in which symptoms are lesser obvious. The issues relating to under-reporting and under-diagnosis are more prevalent in rural areas, where not only diagnostic facilities and trained staff are scarcely available, but cancer registries are also not present [45]. Therefore, the leukemia estimates are reported with wider uncertainty intervals in the case of several LMICs when compared with estimates for HICs. Moreover, leukemia either gets under-reported or misdiagnosed or may not be classified under four main subtypes in LMICs. As a result, the burden of leukemia reflected in the category "other leukemia" is higher in case of low and middle income regions in comparison to high income regions.

Conclusion

Sustainable Development Goal 3 envisages reduction of deaths due to non-communicable diseases (including neoplasms) by one-third by 2030 [54]. However, as per our projections, there would be 720,168 incident cases and 367,804 deaths due to leukemia in 2030 up from 643,579 cases and 334,592 deaths in 2019. Substantial improvements have been witnessed in mortality rates in HICs owing to improved diagnostic and therapeutic capabilities. Early diagnosis, risk stratification, appropriate dose management, and treatment completion are the hallmarks of leukemia management and control in several HICs. Massive progress, however, is required in LMICs, which accounted for 63.7% of total incident cases and 68.6% of deaths due to leukemia in 2019. Moreover, leukemia cases and death estimates are expected to be substantially downward biased in LMICs, and a substantial part of the estimated burden is categorized as other leukemia. Therefore, the actual burden is expected to be much higher in LMICs lacking diagnostic capacity and quality population-based cancer registries. For policy formulation and devising interventions, accurate assessment of leukemia burden and its subtypes is critical in LMICs. For this purpose, diagnostic capabilities and cancer registeration must be boosted in LMICs.

In summary, improving leukemia outcomes in LMICs require tackling challenges relating to diagnosis, availability and cost-effectiveness of therapeutic options, out-of-pocket expenditures, therapy abandonment, deaths due to treatment failure (therapy abandonment, drug toxicity, deaths due to infections), and cancer registration.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s00277-022-04843-6.

Data availability All the data is publicly available on the website http://ghdx.healthdata.org/gbd-results-tool.

Declarations

Ethics approval and consent to participate The research was conducted using data available in the public domain and did not include any human participants or animals. Therefore, no ethical approvals were required.

Consent for publication The authors of the study have consent and responsibility for submission to the journal.

Informed consent Not applicable.

Conflict of interest The authors declare no competing interests.

Permission to reproduce material from other sources All the data used in this study is available in public domain. The geographical maps in the manuscript are drawn using GADM in Python and doesn't reflect author's views on geographical boundaries by the represented countries.

References

1. Ou Z, Yu D, Liang Y et al (2020) Analysis of the Global Burden of Disease study highlights the trends in death and disability-adjusted life years of leukemia from 1990 to 2017. Cancer Commun 40(11):598–610
2. Gupta S, Antillon FA, Bonilla M et al (2011) Treatment-related mortality in children with acute lymphoblastic leukemia in Central America. Cancer 117(20):4788–4795
3. Umeh CA, Garcia-Gonzalez P, Tremblay D, Laing R (2020) The survival of patients enrolled in a global direct-to-patient cancer medicine donation program: The Glivec International Patient Assistance Program (GIPAP). EClinicalMedicine 19:100257
4. Vos T, Lim SS, Abbafati C et al (2020) Global burden of 369 diseases and injuries in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396:1204–1222
5. Wang H, Abbas KM, Abbasifard M et al (2020) Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the global burden of disease study 2019. Lancet 396:1160–1203
6. Murray CJ, Aravkin AY, Zheng P et al (2020) Global burden of 87 risk factors in 204 countries and territories, 1990–2019: a systematic analysis for the global burden of disease study 2019. Lancet 396:1223–1249
7. Global Burden of Disease Collaborative Network. Global Burden of Disease Study 2019 (GBD 2019) Results. Seattle, United States: Institute for Health Metrics and Evaluation (IHME). Source: Institute for Health Metrics Evaluation. Used with permission. All rights reserved. Available from http://ghdx.healthdata.org/gbd-results-tool. Accessed: Jan-Feb 2021
8. https://blogs.worldbank.org/opendata/new-world-bank-country-classifications-income-level-2021-2022. Accessed 23 Nov 2021
9. Jabbour E, Fava C, Kantarjian H (2009) Advances in the biology and therapy of patients with chronic myeloid leukaemia. Best Pract Res Clin Haematol 22(3):395–407
10. Yamamoto JF, Goodman MT (2007) Patterns of leukemia incidence in the United States by subtype and demographic characteristics, 1997–2002. Cancer Causes Control 19(4):379–390
11. Thygesen LC, Nielsen OJ, Johansen C (2009) Trends in adult leukemia incidence and survival in Denmark, 1943–2003. Cancer Causes Control 20(9):1671–1680
12. Berger U, Maywald O, Pfirrmann M et al (2005) Gender aspects in chronic myeloid leukemia: long-term results from randomized studies. Leukemia 19:984–989
13. Höglund M, Sandin F, Simonsson B (2015) Epidemiology of chronic myeloid leukaemia: an update. Ann Hematol 94(Suppl 2):S241-247
14. Bonner MR, Williams BA, Rusiecki JA et al (2010) Occupational exposure to terbufos and the incidence of cancer in the Agricultural Health Study. Cancer Causes Control 21(6):871–877
15. IARC (2021) List of classifications by cancer sites with sufficient or limited evidence in humans, IARC Monographs Volumes 1–129a. Available at https://monographs.iarc.who.int/wp-content/uploads/2019/07/Classifications_by_cancer_site.pdf. Accessed 14.5.2021
16. Poynter JN, Richardson M, Blair CK et al (2016) Obesity over the life course and risk of acute myeloid leukemia and myelodysplastic syndromes. Cancer Epidemiol 40:134–140
17. Weihe P, Spielmann J, Kielstein H et al (2020) Childhood Obesity and Cancer Risk in Adulthood. Curr Obes Rep 9:204–212
18. Shamriz O, Leiba M, Levine H et al (2017) Higher body mass index in 16–19 year-old Jewish Adolescents of North African, Middle Eastern and European Origins is a Predictor of Acute Myeloid Leukemia: a cohort of 2.3 million Israelis. Cancer Causes Control 28:331–339
19. Greaves M (2018) A causal mechanism for childhood acute lymphoblastic leukaemia. Nat Rev Cancer 18(8):471–484
20. Greaves M, Cazzaniga V, Ford A (2021) Can we prevent childhood Leukaemia? Leukemia 35:1258–1264
21. Druker BJ, Guilhot F, O'Brien SG et al (2006) Five-year follow-up of patients receiving imatinib for chronic myeloid leukemia. N Engl J Med 355(23):2408–2417
22. Bower H, Björkholm M, Dickman PW, Höglund M, Lambert PC, Andersson TM (2016) Life Expectancy of Patients With Chronic Myeloid Leukemia Approaches the Life Expectancy of the General Population. J Clin Oncol 34(24):2851–2857
23. Malhotra H, Radich J, Garcia-Gonzalez P (2019) Meeting the needs of CML patients in resource-poor countries. Hematology Am Soc Hematol Educ Program 2019(1):433–442
24. Umeh CA, Garcia-Gonzalez P, Tremblay D, Laing R (2020) The survival of patients enrolled in a global direct-to-patient cancer medicine donation program: The Glivec International Patient Assistance Program (GIPAP). EClinicalMedicine 1(19):100257
25. Allemani C, Matsuda T, Di Carlo V et al (2018) Global surveillance of trends in cancer survival 2000–14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet 391(10125):1023–1075
26. Navarrete M, Rossi E, Brivio E et al (2014) Treatment of childhood acute lymphoblastic leukemia in central America: A lower-middle income countries experience. Pediatr Blood Cancer 61(5):803–809
27. Gupta S, Yeh S, Martiniuk A et al (2013) The magnitude and predictors of abandonment of therapy in paediatric acute leukaemia in middle-income countries: a systematic review and meta-analysis. Eur J Cancer 49(11):2555–2564
28. Sharma R (2021) A Systematic Examination of Burden of Childhood Cancers in 183 Countries: Estimates from GLOBOCAN 2018. Eur J Cancer Care 30(5):e13438
29. Caniza MA, Odio C, Mukkada S et al (2015) Infectious complications in children with acute lymphoblastic leukemia treated in low-middle-income countries. Expert Rev Hematol 8(5):627–645
30. Hunger SP, Lu X, Devidas M et al (2012) Improved survival for children and adolescents with acute lymphoblastic leukemia

between 1990 and 2005: a report from the children's oncology group. J Clin Oncol 30(14):1663–1669
31. Pulte D, Jansen L, Gondos A et al (2014) Survival of adults with acute lymphoblastic leukemia in Germany and the United States. PLoS ONE 9(1):e85554
32 Institute of Medicine (2007) Cancer Control Opportunities in Low- and Middle-Income Countries. The National Academies Press, Washington. DC. https://doi.org/10.17226/11797
33. Liu-Dumlao T, Kantarjian H, Thomas DA, O'Brien S, Ravandi F (2012) Philadelphia-positive acute lymphoblastic leukemia: current treatment options. Curr Oncol Rep 14(5):387–394
34. Maino E, Sancetta R, Viero P et al (2014) Current and future management of Ph/BCR-ABL positive ALL. Expert Rev Anticancer Ther 14(6):723–740
35. Soverini S, Bassan R, Lion T (2019) Treatment and monitoring of Philadelphia chromosome-positive leukemia patients: recent advances and remaining challenges. J Hematol Oncol 12:39
36. Shah A, Andersson TM, Rachet B, Björkholm M, Lambert PC (2013) Survival and cure of acute myeloid leukaemia in E ngland, 1971–2006: a population–based study. Br J Haematol 162(4):509–516
37. Meyers J, Yu Y, Kaye JA, Davis KL (2013) Medicare fee-for-service enrollees with primary acute myeloid leukemia: an analysis of treatment patterns, survival, and healthcare resource utilization and costs. Appl Health Econ Health Policy 11(3):275–286
38. Smith M, Barnett M, Bassan R, Gatta G, Tondini C, Kern W (2004) Adult acute myeloid leukaemia. Crit Rev Oncol Hematol 50(3):197–222
39. Tallman MS, Wang ES, Altman JK et al (2019) Acute myeloid leukemia, version 3.2019, NCCN clinical practice guidelines in oncology. J Natl Compr Canc Netw 17(6):721–49
40. Haferlach C, Dicker F, Schnittger S, Kern W, Haferlach T (2007) Comprehensive genetic characterization of CLL: a study on 506 cases analyzed with chromosome banding analysis, interphase FISH, IgV H status and immunophenotyping. Leukemia 21(12):2442–2451
41. Mrózek K, Eisfeld AK, Kohlschmidt J et al (2019) Complex karyotype in de novo acute myeloid leukemia: typical and atypical subtypes differ molecularly and clinically. Leukemia 33(7):1620–1634
42. de Lira Benicio MT, Ribeiro AF, Américo AD et al (2017) Evaluation of the European Leukemia Net recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low-and middle-income countries (LMIC): A Brazilian experience. Leuk Res 60:109–114
43. Pui CH, Robinson LL, Look AT (2008) Acute myeloid leukemia. Lancet 371:1030–1043
44. Gupta S, Howard SC, Hunger SP et al (2015) Treating childhood cancer in low- and middle-income countries. In: Cancer: Disease Control Priorities, Vol. 3. The International Bank for Reconstruction and Development / The World Bank, Washington, DC
45. Magrath I, Steliarova-Foucher E, Epelman S et al (2013) Improving cancer care for children and young people 2. Lancet Oncol 13:104–106
46. Lam CG, Howard SC, Bouffet E, Pritchard-Jones K (2019) Science and health for all children with cancer. Science 363(6432):1182–1186
47. Philip CC, Geoge B, Ganapule A et al (2014) Acute myeloid leukemia: Challenges and real world data from India. Blood 124(21):3685
48. Lozano R, Garrido F. Improving health system efficiency (Health Systems Governance and Financing). World Health Organization; 20(1). Accessed at https://cdn.who.int/media/docs/default-source/financial-protection/catastrophic-health-expenditure-fund-mexico-case-study.pdf. Last Accessed May 5th, 2021
49. Knaul FM, Gonzalez-Pier E, Gomez-Dantes O et al (2012) The quest for universal health coverage: acheiving social protection for all in Mexico. Lancet 380:1259–1279
50. Antillon F, Baez F, Barrantes Zamorra JC (2005) AMOR: a proposed cooperative effort to improve outcomes of childhood cancer in Central America. Pediatr Blood Cancer 45:107–110
51. Mostert S, Sitaresmi MN, Gundy CM et al (2010) Comparing childhood leukaemia treatment before and after the introduction of a parental education programme in Indonesia. Arch Dis Child 95:20–25
52. Israels T, Borgstein E, Pidini D et al (2012) Managment of children with a Wilms tumor in Malawi, sub-saharan Africa. J Pediatr Hematol Oncol 34:606–610
53. Howard SC, Metzger ML, Wilimas JA et al (2008) Childhood cancer epidemiology in low-income countries. Cancer 112:461–472
54. WHO. Sustainable Development Goal 3. Available at https://www.who.int/health-topics/sustainable-development-goals#tab=tab_2. Accessed 27.5.2021

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

MĀPAN-Journal of Metrology Society of India
https://doi.org/10.1007/s12647-022-00551-8

ORIGINAL PAPER

Measurement of Tribological Characteristics of Composite Coating Under Different Temperatures

K. Gupta*, Q. Murtaza and N. Yuvraj

Department of Mechanical Engineering, Delhi Technological University, Delhi, India

Received: 15 January 2022 / Accepted: 22 March 2022

© Metrology Society of India 2022

Abstract: In the present work, the measurements of tribological properties of Mo–NiCr–WC–W blend composite coating on the mild steel substrate deposited using high velocity oxy-fuel thermal spray technique under various temperatures have been presented. Initially, the powders for coating were mixed employing a laboratory ball mill. The composite coating is done via a well-known high velocity oxy-fuel thermal spray technique. Further, different test specimens were prepared to measure the tribological behaviour and residual stress. High-temperature pin-on-disc tribometer was used to investigate the wear and friction with increasing temperature, load and sliding speed. X-ray diffraction, field emission scanning electron microscope, and micro-hardness tester were used to characterize the coating, wear track and micro-hardness of the coatings, respectively. The weight-loss method and machine output were used to analyze the coefficient of friction and specific wear rate. Based on the said test, responses have been obtained, analyzed, and discussed in detail.

Keywords: Measurement; Temperature; Mo–NiCr–WC–W composite coating; HVOF thermal spray coating; FESEM

1. Introduction

Metals and metallic alloys are always been in demand since long back to meet the specific industrial requirements of automobile, aerospace, aircraft defence and many more industries. However, it is quite challenging to satisfy the various specific requirements by the single material as the serving conditions are obviously different for different industries' products [1, 2]. In general, the structural [3–7] as well as tribological behaviours [8–10] of the material such as hardness, resistance to erosion and resistance to abrasion are considered as important parameters in the selection of the material for various specific applications. However, it is also true that the manufacturing of industrial products or different parts of the products with the aforementioned properties is quite difficult and not always practical and economically feasible too. The metal surface coating [11] can be considered as one of the promising, feasible and economical solutions to such problems. The surface coating is a technology in which a very thin layer of metal or the combination of various metals also called composite is deposited over the surface of another metal to improve the surface quality/properties (tribological) as per the requirement. Molybdenum (Mo), chromium (Cr), tungsten (W), tungsten carbide (WC), nickel–chromium (NiCr), chromium carbide (CrC), silicon carbide (SiC), etc., or a composite of all these powders [12–14] can be used for the same. The hard-metal coating of materials like W, Cr and WC often provides the required hardness, toughness, high wear and corrosion resistance [13–15] to the base metals. Whereas Mo particles are often used for the required improvement in melting point, coefficient of thermal expansion, strength and durability. Also, the Mo application as a coating material reduces friction by a notable amount [16].

The powdered blend of Mo, NiCr, W, WC, CrC, SiC, etc., can be deposited on the substrate with the use of the high velocity oxy-fuel coating (HVOF) technique [17–19]. Various researchers have reported [11, 20] that the hard-metal coatings deposited by the HVOF thermal spray method exhibit a high density of deposited particles, low porosity and good mechanical strength. This is due to the particularly high momentum of the feedstock particles at the instant of their impact with the selected substrate which results in a high degree of adhesion of the powdered mixture on the substrate. HVOF thermal sprayed coatings are also considered to be more environmentally friendly and better-performing coating alternatives to electroplated hard chromium [16, 20].

*Corresponding author, E-mail: kalpanaguptapandey@gmail.com

Published online: 03 May 2022

From the literature review, it is found that the tribological characteristics of the metal can be improved by the composite coating. It is also found that HVOF is an efficient technique that can be used for composite coating. Additionally, the elements which are individually responsible for the various properties improvement can be used together to improve the surface quality. In view of the above, multiple substrates of mild steel have been coated with a blend of W–WC–NiCr–Mo composite coating in the present work for tribological applications. The effect of temperature and load on the tribological properties as well as residual stress of the W–WC–NiCr–Mo blend composite coatings has been investigated. Further, the morphology of coating, microstructure, the particle size distribution of the powder, wear mechanism, and damping resistance of the coating have also been analyzed and discussed in detail.

2. Materials and Methods

The experimental procedure involves three steps. The first step involves the preparation of powder and substrate. The second step involves the preparation of coating. In the third step, tribological investigation has been done.

The powders of Mo, NiCr, W, and WC with average sizes varying from 40 to 65 μm ± 5 μm are used for the feed preparation. Scanning electron microscope (SEM) micrographs of powder can be seen in Fig. 1a-d. First of all, the powders are mixed together using a laboratory ball mill for 2 h. After ensuring the uniform mixing, the mixture is then heated at 100 °C for 1 h to remove the moisture before loading into the feeder. Now, the substrates are sandblasted using SiC powder with 45 μm grit size and MEC make sandblasting machine. The sandblasting is done to increase the roughness so that the mechanical anchorage of the deposited powder can be increased. In this work, a mild steel substrate material with a dimension of 80 mm × 80 mm × 2.5 mm is used for the development of the W–WC–NiCr–Mo composite coating. The chemical composition of the said substrate is highlighted in Table 1. The HVOF technique is then used for the deposition of the composite powder on the low carbon steel substrate. Also, the schematic diagram of the HVOF coating setup is shown in Fig. 2. The process parameters of the HVOF machine are reported in Table 2.

The LPG liquefies while spraying and this liquid interfered with the flow metre during the experimentation because of which we were unable to get the accurate reading for the fuel gas. To stop liquid interference, it is suggested to use a good vaporizer.

3. Experimental Investigation

A pin-on-disc high-temperature tribometer is used for the measurement of tribological characteristics of the W–WC–NiCr–Mo composite coating. A schematic diagram of the tribo-test machine is portrayed in Fig. 3. The operational attachment such as advance motion control module, load (20–200 N), sliding velocity (up to 3000 rpm), temperature control (850 °C and 800 °C for the disc and pin, respectively) and linear variable differential transformer (LVDT) is used for the wear measurement. Heating of the pin and disc both can be done, but in this research, the temperature

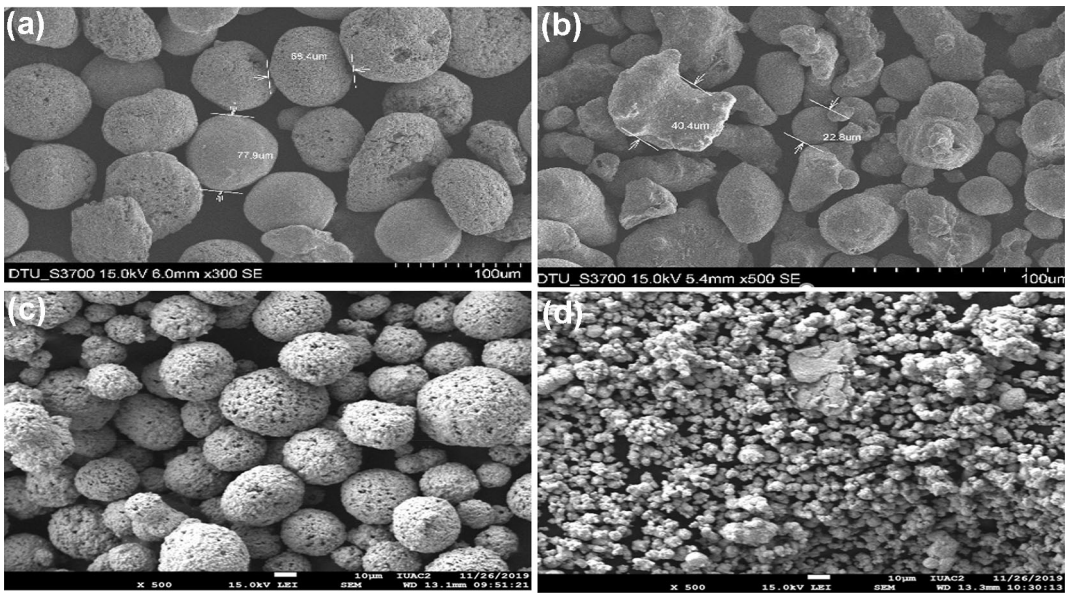


Fig. 1 SEM micrographs of feed powders **a** Mo powder **b** NiCr powder **c** W powder **d** WC powder

Table 1 Chemical composition of mild steel substrate

Element	Carbon (C)	(Manganese) Mn	Silicon (Si)	Phosphorus (P)	Molybdenum (Mo)	Aluminium (Al)	Copper (Cu)
Composition (%)	0.15	0.11	0.19	0.015	0.02	0.05	0.08

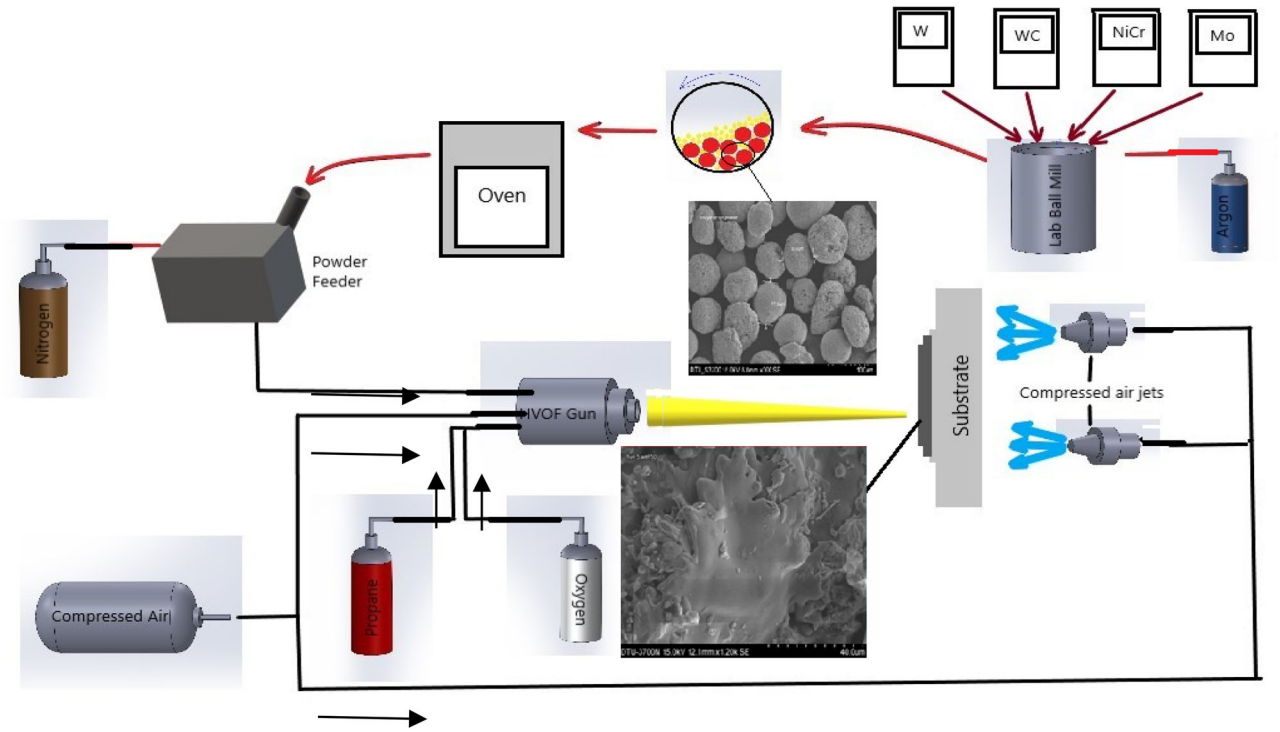


Fig. 2 Schematic diagram of the HVOF coating setup

Table 2 Process parameter of HVOF coating for the deposition

S. No	Parameters	Values
1	Fuel (LPG) pressure kg/cm^2	06
2	Oxygen pressure kg/cm^2	08
3	Air pressure kg/cm^2	05
4	Carrier gases pressure (N_2) kg/cm^2	5
5	Oxygen flow (SLPM)	250
6	Fuel flow (SLPM)	50
7	Air flow (SLPM)	100
8	Carrier gas flow (SLPM)	15

of the disc only is increased. The wear test is performed for four sets, and during every tribo-test, a pin of Ø8 mm of EN-31 and the disc (composition in Table 1) of Ø60 mm and thickness 2 mm are used. The pin is used as a counter body, and the contact shape between pin and disc is flat circular of Ø8 mm. Finally, the tribological properties of the developed composite coating are studied for the parameters such as load (N), temperature (°C), sliding velocity (m/s) and sliding distance (m). The various tribo-test conditions used while performing tests are reported in Table 3, 4, 5.

3.1. Microstructure Characterization of Composite Coating

In order to perform the microstructure characterization of the composite coating (W–WC–NiCr–Mo), energy-dispersive X-ray spectroscopy (EDS), field emission scanning electron microscope (FESEM) and X-ray diffraction (XRD) techniques are used. X'Pert High Score Plus software was used to analyze the phase present in the deposited composite coating. The atomic and molecular arrangements in the composite coating were studied with the help of the diffraction phenomena. The X-ray beams incident on the crystalline material may undergo constructive interference because of its contact with a series of parallel atomic planes conferring to Bragg's law. The Miller indices and lattice parameter(s), as well as crystal structure, influence interplanar spacing. A residual stress analyzer μ-X360 was used to assess the residual stresses that formed on the coated surfaces via the cosine α method. The

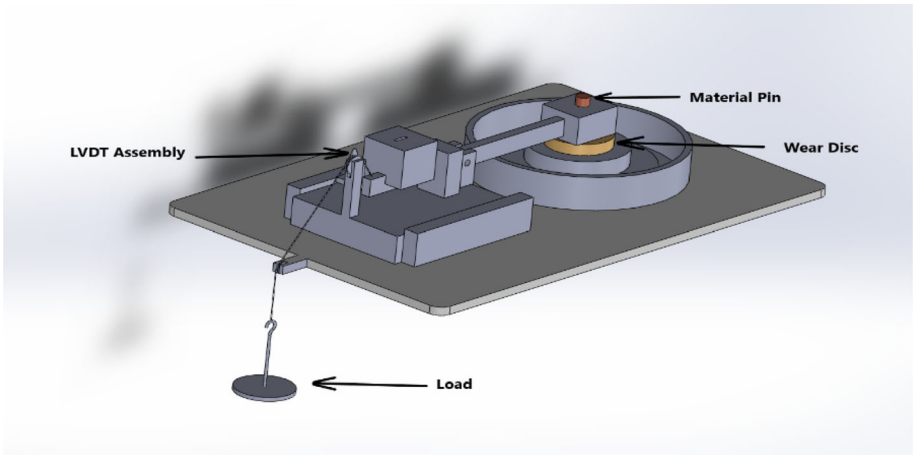


Fig. 3 Schematic diagram of pin-on-disc tribometer

Table 3 Tribo-test condition

Sr. No	Parameters	Values
1	Temperature	100 °C, 200 °C, 300 °C, 400 °C
2	Load	55 N, 65 N, 75 N, 85 N
3	Speed	1.0 m/s, 1.5 m/s, 2.0 m/s, 2.5 m/s

Table 4 Experimental condition

Test No	Temperature (°C)	Load (N)	Speed (m/s)
Test 01	100	55	1.0
Test 02	200	65	1.5
Test 03	300	75	2.0
Test 04	400	85	2.5

analyzer had a power and sensor unit, a light-emitting diode (LED) marker, and charged coupled device (CCD) camera support that is accessible at the precision manufacturing laboratory, Delhi Technological University, New Delhi. The mathematical formulation and graphical representation of the cosine-α method are illustrated in Fig. 4 and Eqs. (1)–(4).

$$\varepsilon\alpha_1 = \frac{1}{2}\{(\varepsilon\alpha - \varepsilon\pi + \alpha) + (\varepsilon - \alpha - \varepsilon\pi - \alpha)\} \quad (1)$$

$$\varepsilon\alpha_2 = \frac{1}{2}\{(\varepsilon\alpha - \varepsilon\pi + \alpha) - (\varepsilon - \alpha - \varepsilon\pi - \alpha)\} \quad (2)$$

$$\sigma_x = \frac{E}{(1+\nu)} \times \frac{1}{(\sin 2\eta)} \times \frac{1}{(\sin 2\Psi_0)} \times \frac{\partial\varepsilon\alpha_1}{(\partial \cos\alpha)} \quad (3)$$

$$\sigma_y = \frac{E}{2(1+\nu)} \times \frac{1}{(\sin 2\eta)} \times \frac{1}{(\sin 2\Psi_0)} \times \frac{\partial\varepsilon\alpha_2}{(\partial \sin\alpha)} \quad (4)$$

3.2. Micro-Hardness Measurement

Examination of the micro-hardness gives the resistance to plastic deformation of the material. During this test, a diamond indenter with a defined geometry is pressed into the test specimen's surface with a specified mechanical force. The formed geometry is assessed for the micro-hardness values. For testing of the coating Fischer micro-hardness tester was used. The indenter was made to be in the contact with the surface mechanically making a negligible impact on the test surface. The full load was applied and maintained for the selected dwell time after that indenter has been withdrawn. After removal of the load, the length of the larger diagonal has been measured with the help of a micrometre eyepiece and the hardness corresponding to that diagonal length has been measured.

3.3. Hardness Measurement

The Vickers micro-hardness was measured using a Fischer make micro-hardness evaluator accessible at Delhi Technological University, New Delhi. The hardness test was

Table 5 Constant parameters

Parameter	Condition
Sliding distance	Sliding distance 1500 m was taken constantly for all the test condition
Counter body	Pin of Ø8mm EN-31 was similar for all the test
Disc	Disc of Ø60mm and thickness 2 mm of mild steel with composite coating W-WC-Mo-NiCr
Lubrication condition	Dry condition

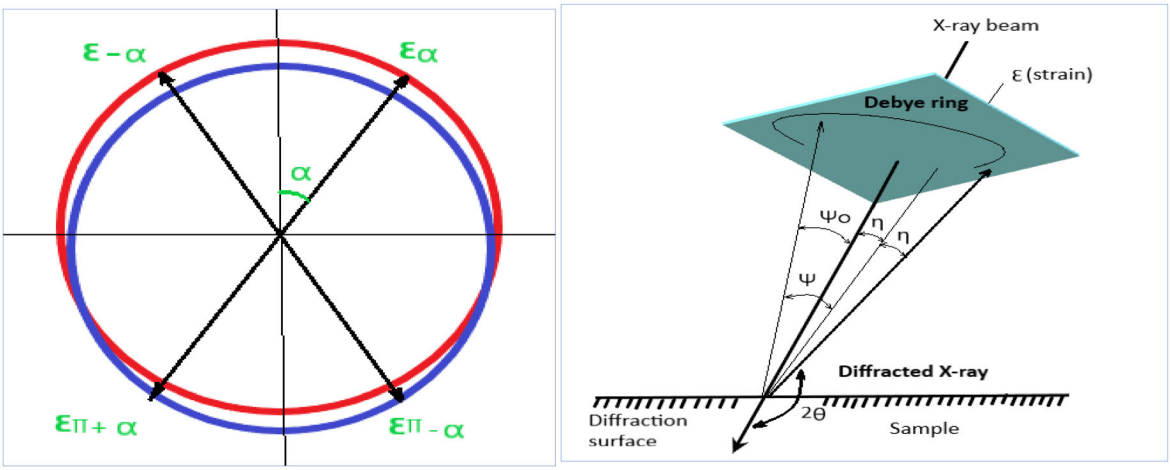


Fig. 4 Graphical representation of cosine-α method

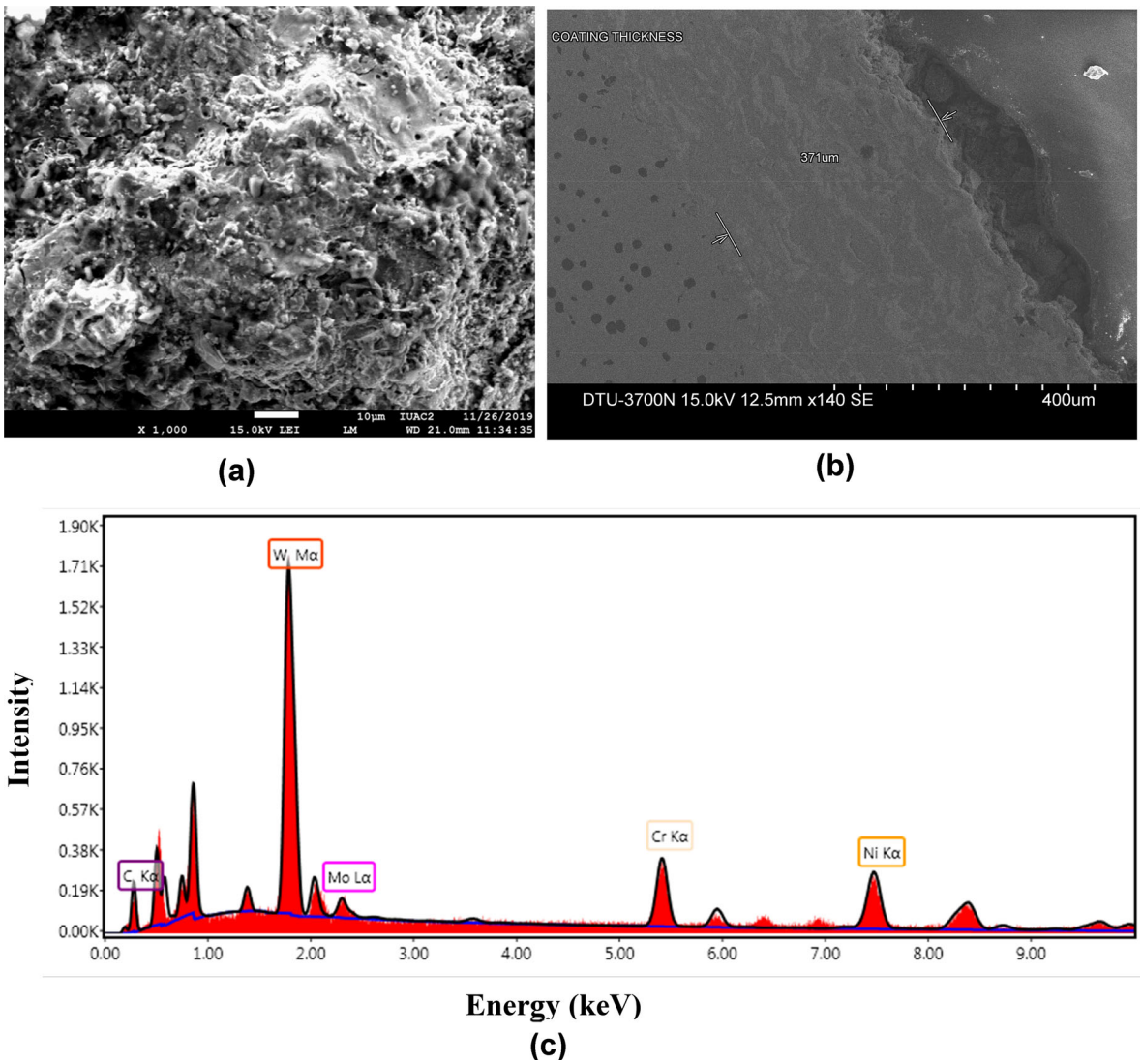


Fig. 5 **a** Surface morphology of the deposited coating **b** coating cross section for the thickness **c** EDX analysis of deposited surface

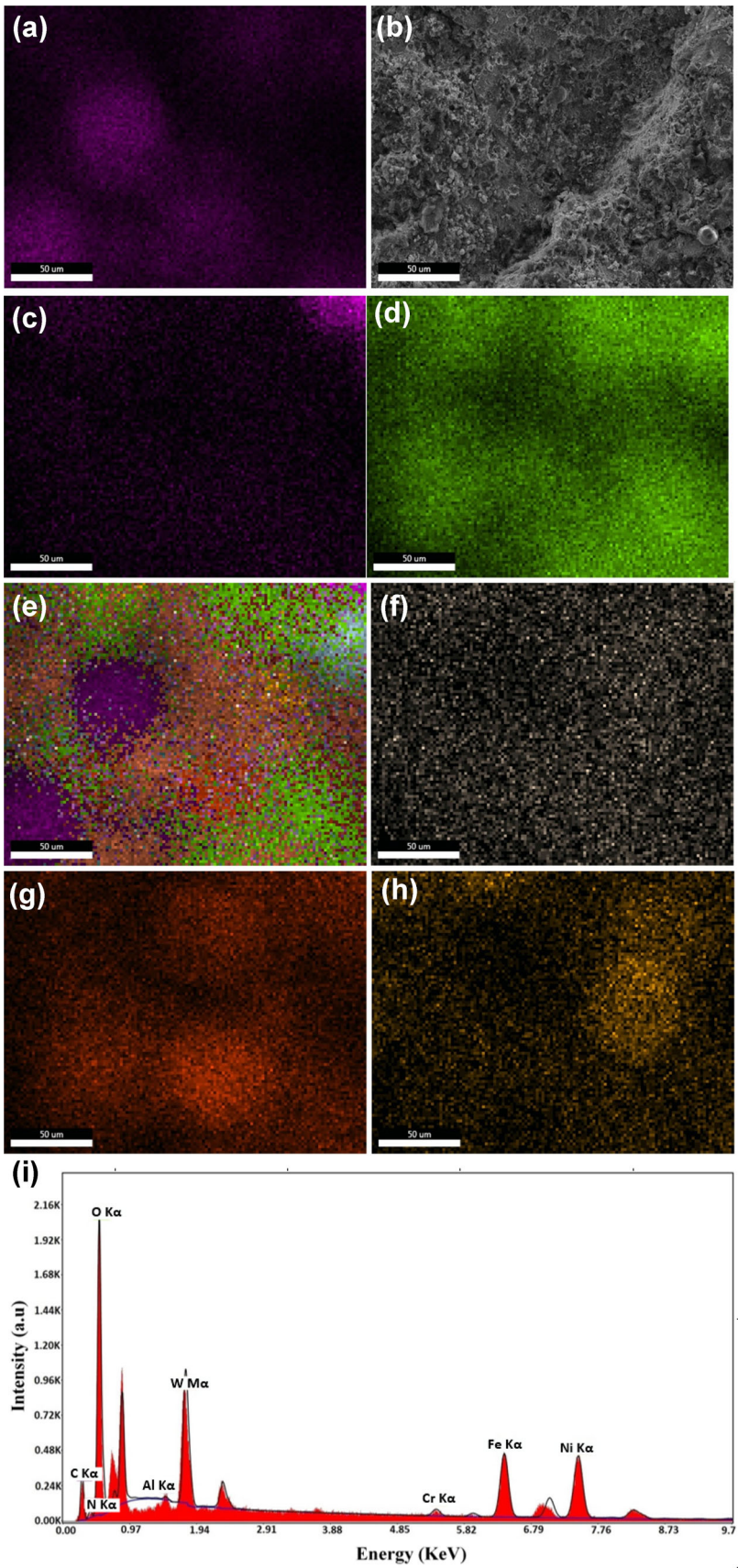


Fig. 6 X-ray image of **a** carbon **b** coating **c** molybdenum **d** oxygen **e** complete mapping **f** chromium **g** tungsten **h** EDX **i** nickel

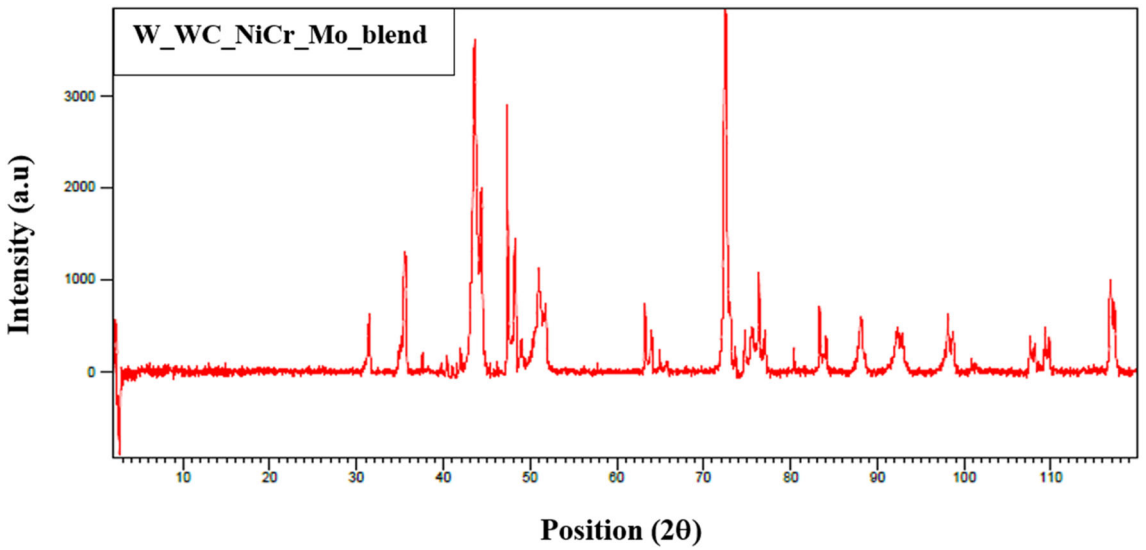


Fig. 7 XRD analysis of blended powder used for the deposition of coating

performed using a force of 300mN (3.05915×10^{-2}Kgf), a 20-s indentation period, and a 5-s dwell time.

3.4. Specific Wear Rate Measurement

The weight loss approach was used to calculate the composite coating's wear. The weight loss was calculated using a weighing machine with the least count of 0.00001 g. The test sample was cleaned with acetone and weighed before and after the test. The specific wear rate (SWR) is calculated using Eq. (5).

$$\mathrm{SWR} = \frac{\mathrm{Volume(mm^3)}}{\mathrm{Sliding\ Distance(m)} \times \mathrm{Applied\ Load(N)}} \tag{5}$$

4. Results and Discussion

4.1. Coating Characterisation

The surface morphology of the deposited composite coating was analyzed with the help of FESEM, XRD and energy-dispersive X-ray spectroscopy (EDX). The deposited coating was in a perfectly molten state. Mostly pancake type of splats was observed on the deposited coating (Fig. 5a, b). The presence of Mo, Ni, Cr and W was also verified with the EDX analysis. It was also observed that Mo, Ni and Cr were present in their oxide form on the deposited surface. In the EDX diagram peaks of Ni, Cr and C are given off as X-ray returns to K electron shell. The peak of tungsten is given off my M electron shell, and the peak of molybdenum is given by the L electron shell (Fig. 5c). The thickness of the coating was found to be 370–380 µm. The layered deposition of the coating is

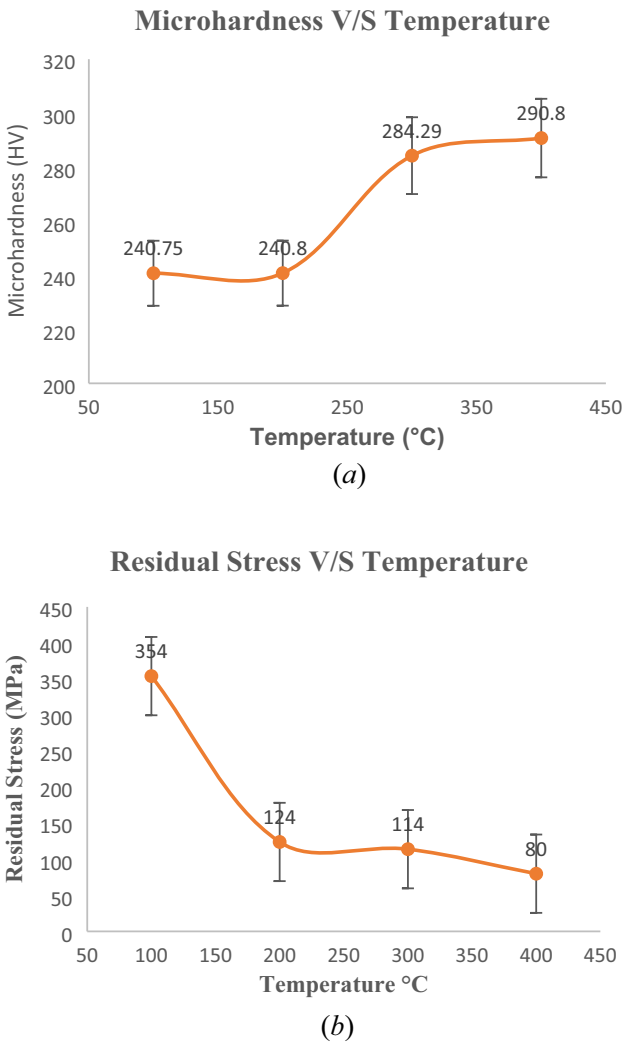


Fig. 8 **a** Micro-hardness versus temperature curve **b** residual stress versus temperature curve

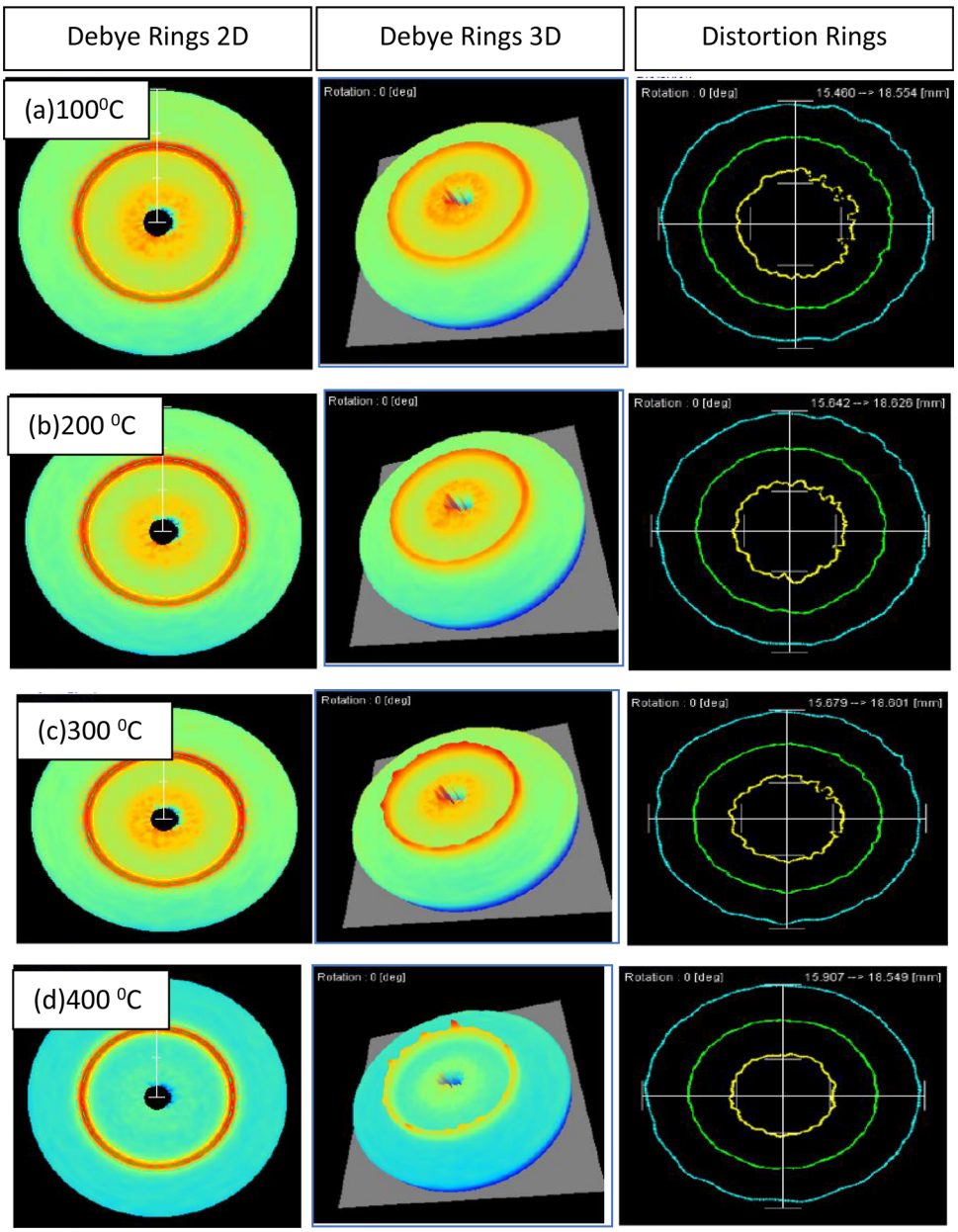


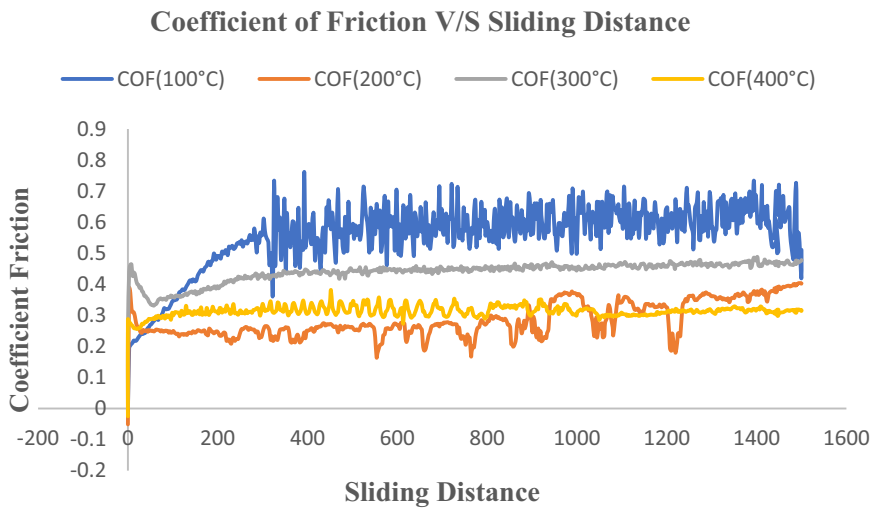
Fig. 9 Debye rings at temperatures **a** and 100, **b** 200, **c** 300 and **d** 400 °C

clearly visible from the cross section of the coating. The coating is found to be uniform throughout the surface of the substrate. Higher magnification shows that matrixes are perfectly bonded with each other which may result in the enhancement of hardness and wear resistance properties of the developed composite coating.

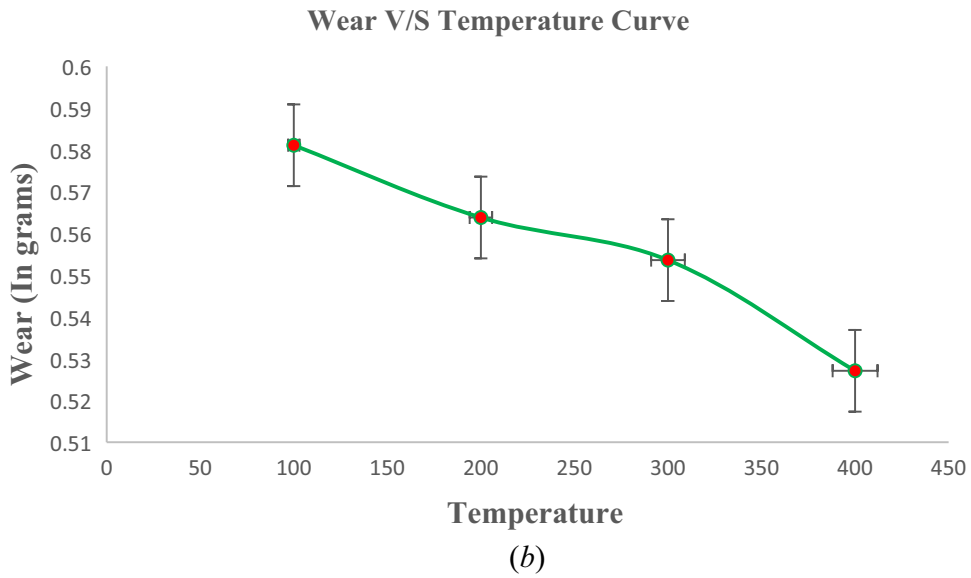
The distribution of the elements in the developed composite coating can be affirmed with the help of elementary mapping of the coating, and the results are shown in Fig. 6. In the figure, carbon elements are shown in the magenta regions in the mapping, whereas purple, green, grey, orange and yellow are respective colours for molybdenum, oxygen, chromium, tungsten and nickel. However, the presence of these elements can also be confirmed with the peaks of EDX images (Fig. 7).

4.2. Micro-Hardness and Residual Stresses Measurement

The micro-hardness of the tungsten-based composite coating was calculated as 240.75 ± 5, 240.8 ± 5, 284.29 ± 5 and 290.8 ± 5 HV at an increasing temperature from 100 to 400 °C in the interval of 100 °C, respectively (Fig. 8a). For tungsten-based composite coating, the micro-hardness value at 400 °C was higher than 100 °C, 200 °C and 300 °C because of interfacial strength amongst the microparticles used in coating and also due to



(*a*)



(*b*)

Fig. 10 **a** Coefficient of friction versus sliding distance **b** wear versus temperature curve

modification of grain positioning at elevated temperature. The coated micro-carbon particles serve as a resistance to plastic deformation and lead to a rise in hardness.

The residual stress (Fig. 8b) of the tungsten-based composite coating was observed as 354 ± 5, 124 ± 5, 114 ± 5 and 80 ± 5 MPa at an increasing temperature from 100 to 400 °C in the interval of 100 °C, respectively. Due to stress relief or change grain orientation of coated material, the residual stress decreases. Figure 9 shows the Debye rings 2D, 3D and distortion ring at various temperatures from 100 to 400 °C in the interval of 100 °C.

4.3. Effect of Load on Wear Measurement

Figure 10a depicts the changes in coefficient of friction of tungsten-based composite coatings at various temperatures, sliding velocity at 1.0 m/s, 1.5 m/s, 2.0 m/s and 2.5 m/s and at load at 55 N,65 N, 75 N and 85 N in dry condition with 1500 m sliding distance. The average value of coefficient of friction from TEST 01, TEST 02, TEST 03 and TEST 04 is 0.622,0.317,0.463 and 0.314 at variable temperature (100 °C, 200 °C, 300 °C, and 400 °C), respectively.

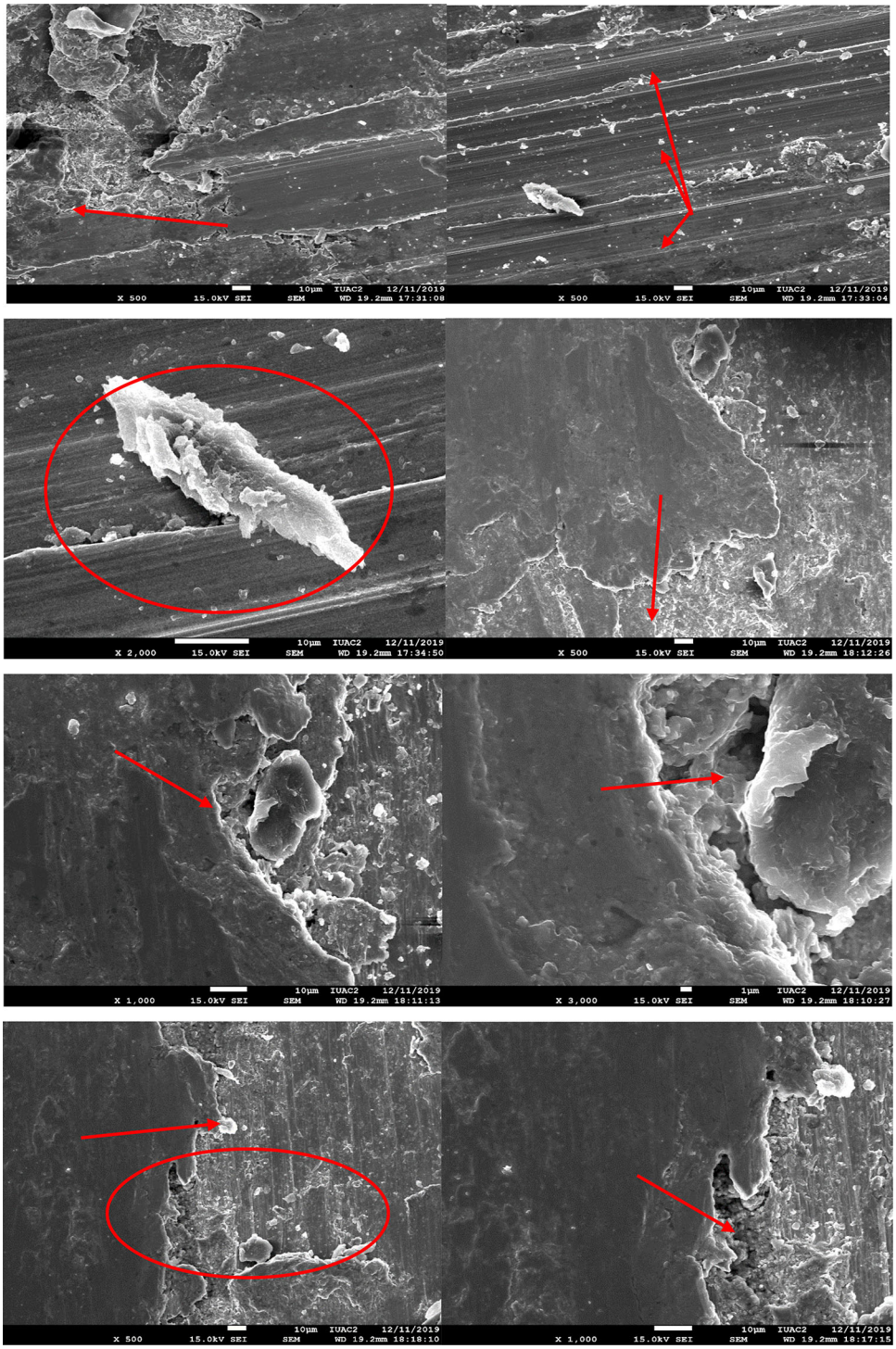


Fig. 11 **a–d** FESEM micrographs of wear out surfaces

4.4. Wear Mechanism

Figure 10b demonstrates the deviation in wear with reference to the change in temperature. The average wear values of tungsten-based composite coating obtained from TEST 01, TEST 02, TEST 03 and TEST 04 are 0.5811, 0.5638, 0.5536, 0.5271 at variable temperature 100 °C, 200 °C, 300 °C, and 400 °C, respectively. The values of wear show the decreasing trend with the increase in temperature, as expected.

To investigate the wear-out behaviour of the composite coating, the FESEM images worn surfaces of tungsten-based composite coating are displayed Fig. 11a–d. The figure assures the existence of abrasive wear. The wear mechanism of wear-out surfaces, micron-size carbon particles, cracks, oxide development, formation and layer debonding verifies erosion and signifies abrasive wear as the prevailing wear phenomenon on the produced coating.

5. Conclusions

The current article presents the measurement of tribological attributes, and residual stress of Mo–NiCr–WC–W blend composite coating on the mild steel substrate deposited employing HVOF thermal spray technique. The wear and friction were investigated at increasing temperature utilizing a pin-on-disc tribometer, and the following inference can be observed through the present work.

- The micro-hardness measurement of the tungsten-based composite coating indicates the increasing trend of hardness with an increase in temperature. The maximum hardness was obtained at the temperature of 400 °C about 290.8Hv.
- The residual stress measurement shows a decreasing behaviour of stress with the increase in temperature. Maximum residual stress was found at 100 °C (354 MPa) and minimum at 400 °C (80 MPa).
- Debye rings also reveal the residual stress at various temperatures from 100 °C to 400 °C.
- Coefficient of friction obtained maximum at 100 °C and minimum at 400 °C with 1500 m sliding distance while at 200 °C and 300 °C were lies between above temperature.
- The wear rate of tungsten-based composite coating shows maximum at 100 °C and minimum at 400 °C.
- FESEM reveals the worn-out surface of tungsten-based composite coating like cracks, worn surfaces, oxide layers, abrasion and delamination of layers.

References

[1] F. Wang, L. Zheng, Q. Li, F. Zhang, X. Chen and H. Zhang, Corrosion Properties of Carbon Ions Implanted Chromium Coating Prepared on CSS-42L Aerospace Bearing Steel. Surf. Coat. Technol., *349* (2019) 392–399.

[2] D.G. Bhosale, W.S. Rathod and S.W. Rukhande, Sliding Wear Behavior of High Velocity Oxy-Fuel Sprayed WC-Cr3C2-Ni Coating for Automotive Applications. Mater. Today Proc., *19* (2019) 1–5.

[3] C.K. Hirwani, S.K. Panda, S.S. Mahapatra, S.K. Mandal, L. Srivastava and M.K. Buragohain, Flexural Strength of delaminated Composite Plate–An Experimental Validation. Int. J. Damage Mech., *27* (2018) 296–329.

[4] C.K. Hirwani, S.K. Panda, T.R. Mahapatra and S.S. Mahapatra, Nonlinear Transient Finite-Element Analysis of Delaminated Composite Shallow Shell Panels. Aiaa J., *55* (2017) 1734–1748.

[5] M.C. Kunche, P.K. Mishra, H.B. Nallal, C.K. Hirwani, P.V. Katariya, S. Panda and S.K. Panda, Theoretical and Experimental Modal Responses of Adhesive Bonded T-Joints. Wind Struct., *29* (2019) 361–369.

[6] H.K. Pandey, C.K. Hirwani, N. Sharma, P.V. Katariya, H.C. Devangan and S.K. Panda, Effect of Nano Glass Cenosphere Filler on Hybrid Composite Eigenfrequency Responses-An FEM Approach and Experimental Verification. Adv. Nano Res., *7* (2019) 419–429.

[7] C.K. Hirwani, P.K. Mishra and S.K. Panda, Nonlinear Steady-State Responses of Weakly Bonded Composite Shell Structure Under Hygro-Thermo-Mechanical Loading. Comp. Struct., *265* (2021) 113768.

[8] S.C. Tung and H. Gao, Tribological Characteristics and Surface Interaction Between Piston Ring Coatings and a Blend of Energy-Conserving Oils and Ethanol Fuels. Wear, *255* (2003) 1276–1285.

[9] R. Soleimani, F. Mahboubi, M. Kazemi and S.Y. Arman, Corrosion and Tribological Behaviour of Electroless Ni-P/Nano-SiC Composite Coating on Aluminium 6061. Surf. Eng., *31* (2015) 714–721.

[10] M. Kalin, E. Roman and J. Vizintin, The Effect of Temperature on the Tribological Mechanisms and Reactivity of Hydrogenated, Amorphous Diamond-Like Carbon Coatings Under Oil-Lubricated Conditions. Thin Solid Films, *515* (2007) 3644–3652.

[11] J.A. Picas, A. Forn and G. Matthäus, HVOF Coatings as an Alternative to Hard Chrome for Pistons and Valves. Wear, *261* (2006) 477–484.

[12] A. Tyagi, S.M. Pandey, R.S. Walia and Q. Murtaza, Characterization and Parametric Optimization of Tribological Properties of Mo Blend Composite Coating. Mater. Res. Exp., *6* (2019) 086428.

[13] G. Bolelli, M. Berger, T. Börner, H. Koivuluoto, L. Lusvarghi, C. Lyphout, N. Markocsan, V. Matikainen, P. Nylén, P. Sassatelli, R. Trache and P. Vuoristoc, Tribology of HVOF- and HVAF-Sprayed WC-10Co4Cr Hardmetal Coatings: A Comparative Assessment". Surf. Coat. Technol., *265* (2015) 125–144.

[14] C.W. Lee, J.H. Han, J. Yoon, M.C. Shin and S.I. Kwun, A Study on Powder Mixing for High Fracture Toughness and Wear Resistance of WC–Co–Cr Coatings Sprayed by HVOF. Surf. Coat. Technol., *204* (2010) 2223–2229.

[15] R. Keshavamurthy, J.M. Sudhan, A. Kumar, V. Ranjan, P. Singh and A. Singh, Wear Behaviour of Hard Chrome and Tungsten Carbide-HVOF Coatings. Mater. Today Proc., *5* (2018) 24587–24594.

[16] C.S. Ramesh, H. Adarsha, A. Chaturvedi and N. Nair, Investigations on the Effect of Molybdenum (Mo) and Molybdenum Silicon Carbide (Mo-10% Sic) Composite Coatings on Mild Steel Substrate Using HVOF Technique. Mater. Today Proc., *5* (2018) 24422–24427.

[17] A. Kiatisereekul and K. Fakpan, High Temperature Erosion High Temperature Erosion of T22 Steel Coated by Cr–Ni Electroplating, HVOF Sprayed Stellite–6 and WC–12Co. Mater. Today Proc. (2021).

[18] Y. Zhou, J. Kang, J. Zhang, Z. Fu, L. Zhu, D. She and W. Yue, Microstructure and Sliding Wear Behavior of HVOF Sprayed Al(1–x)CoCrFeNiTix High-Entropy Alloy Coatings. Mater. Lett., *314* (2022) 131929.

[19] B. Somansundaram, V.G. Patil, M.R. Ramesh, S. Kandaiah, and N. Jegadeeswaran, Microstructure Characteristics and Properties

of WC–CrC–Ni HVOF Coating for Boiler Tube Steel. Mater. Today Proc. (2021).
[20] J.A. Picas, M. Punset, M. Teresa Baile, E. Martín and A. Forn, Tribological Evaluation of HVOF Thermal-Spray Coatings as a Hard Chrome Replacement. Surf. Interf. Anal., *43* (2011) 1346–1353.

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360084536

Modelling and Inverse Kinematics of Tendon Driven Continuum Robots using Cascade Correlation

Article *in* International Journal of Engineering and Technical Research · April 2022

CITATIONS
0

READS
54

3 authors, including:



Shaswat Garg
Delhi Technological University
2 PUBLICATIONS **0** CITATIONS
SEE PROFILE



Satwik Dudeja
Delhi Technological University
1 PUBLICATION **0** CITATIONS
SEE PROFILE

All content following this page was uploaded by Shaswat Garg on 21 April 2022.

The user has requested enhancement of the downloaded file.

Published by :
http://www.ijert.org

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

Modelling and Inverse Kinematics of Tendon Driven Continuum Robots using Cascade Correlation

Ashok Madan
Mechanical Engineering Department
Delhi Technological University
New Delhi, India

Shaswat Garg
Mechanical Engineering Department
Delhi Technological University
New Delhi, India

Satwik Dudeja
Mechanical Engineering Department
Delhi Technological University
New Delhi, India

***Abstract*— Inverse kinematics is among the most important challenges in robotics, since it is required to map the design space onto the joint space in order to permit smooth motion throughout any task. Inverse Kinematics is difficult to master because of the non-linearities and kinematic redundancy seen in continuum robots. We offer a static model for a tendon-driven continuum robot in this study, which is based on Cosserat string theory and rod theory. For inverse kinematics, we suggest using a neural network based on cascade correlation. Using 20,000 established training data, Cascade Correlation model reaches RMSE of 1.34 and 1.45 in training and testing. Comparative study is done by replicating models from previous literature, validating the accuracy of our proposed model.**

Keywords – Inverse Kinematics, Cascade Correlation, Continuum Robots, Tendon Driven Robots

I. INTRODUCTION

Since the last decade, continuum robots [1] seem to be a prominent problem in the context of robotics. For the researchers, nature has been a great source of inspiration. The remarkable mobility, manipulation, and dexterity skills of elephant trunks [2], snakes [3], and octopuses [4] in an unpredictable and chaotic environment have prompted researchers to construct continuum robots. A continuum robot is a robot with limitless degrees of freedom that bends continuously. In comparison to traditional robots, continuum machines can perform in complicated obstacle fields with relative ease due to their innate agility. They are commonly utilized for surgical operations [5] and disaster response in collapsed structures [6] due to their compliance. Pneumatic actuators [7], concentric tubes [8], tendon driven mechanisms [10], and shape memory alloys [9] can all be used to control continuum robots.

The kinematics of continuum robots is a fascinating and difficult topic to solve. Kinematics may be broken down into two types of mappings. From actuator or joint space, to configuration space is the first step. Aside from configuration space, there is task space, which describes the position and direction of the backbone. Two common strategies for tackling the problem are model-based [11] and machine learning [12]. Low fidelity lumped parameter models and high fidelity distributed parameter models are the two types of continuous robot modelling efforts [13]. External loadings like gravity cause variations in curvature over the length of the robot, which low-fidelity models disregard. To tackle kinematics problems, Jones B. et al. [14] designed a DH parameter representation. The kinematics of the robot were approximated as a set of stiff links connected by revolute and prismatic joints. To develop the kinematic equations, Wang He. et al. [15] suggested an analytical technique based on the Euler-Bernoulli beam. Arc geometry was utilized by Li M. et al. [16] to establish the kinematic relationship across the configuration space and the task space. The performance of lumped parameter models in real time is outstanding, but accuracy suffers. As a result, distributed parameter high fidelity models are employed to simulate an arbitrary form of a continuum robot. Jones B. et al. [17] proposed utilizing Cosserat rod theory to compute the form of a continuum robot. Rucker D. et al. [18] built on prior research by taking into account attachment point force, and distributed wrench tendon applications on the backbone. S. Song et al. [19] suggested a reconstruction approach based on a third-order Bezier curve for positional and directional data acquired from electromagnetic sensors positioned on the flexible robot's distal end. To develop the kinematics of a continuum robot, Ding J. et al. [20] introduced a pseudo rigid body model of beam.

Despite their excellent accuracy, distributed parameter models are computationally costly to solve. As a result, machine learning algorithms are employed since they are very accurate and can perform in real time once learned. Multilayer perceptron was utilized by Kuntz A. et al. [21] to anticipate the form of a 3-tube concentric tube robot. Using RBF, MLP, SVR, and CANFIS, Loutfi I. et al [22] calculated the forward kinematics of Compact Bionic Handling Assistant. A backpropagation neural network was used by Shahabi E. et al. [23] to train the inverse kinematics of a dual backbone robot. The dataset was created using the PRBM 3R model. G. Wu et al. [24] used RBF, MLP, and a combination of Genetic algorithm and MLP to study the kinematics of a parallel continuum robot. Lai et al. [25] used neural networks to address the problem for a dual segment continuum robot. DH Parameter Formulation was used to create the dataset. Shen W. et al. [26] used Gaussian Process Regression to tackle the forward kinematics problem of a Tendon-driven continuum robot. Inverse kinematics of a tendon-driven serpentine robot have also been devised using algorithms such as KNNR and Extreme learning machine

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

[27]. Thuruthel et al. [28] developed an MLP model to deal with several solution problems in inverse kinematics utilizing diverse mappings across configuration and task space.

Many existing learning-based kinematic solutions for continuum robots, on the other hand, have a large degree of inaccuracy, making them unsuitable for application in contexts such as surgical operations, where precision is critical. As a result, we concentrate on employing invertible neural networks to solve the problem. The method is evaluated on a Cosserat Rod theory-based Tendon-driven continuum robot with DOFs.

The following is how the rest of the paper is organised: The design of the Tendon-driven continuum robot is presented in Section II. The forward kinematic modelling study utilising Cosserat Rod theory is presented in Section III. In Section IV, a cascade correlation network model is used to train the robot's inverse kinematics. The results of the suggested integrated accuracy enhancement approach are presented in Section V to validate its efficacy. The conclusion of this study is presented in Section VI.

II. MODELLING OF CONTINUUM ROBOT

The Cosserat model introduced in [29] is used to represent the kinematics of the continuum robot. Both the tendon attachment points force and the distributional wrench the tendon exerts along the length of the backbone are taken into account in the aforementioned model. The backbone is defined by a p(s) center-line curve and R orientation (s). Therefore, the homogenous rigid body transformation is given by –

$$g(s) = \begin{bmatrix} R(s) & p(s) \\ 0 & 0 \end{bmatrix}$$

Given the linear and angular rates of change of $g(s)$ as $v(s)$ and $u(s)$. Thus, the evolution of $g(s)$ along s is defined as –

$$\dot{R}(s) = R(s)\hat{u}(s)$$
$$\dot{p}(s) = R(s)v(s)$$

The equations of static equilibrium of an arbitrary section of rod are given by [30] –

$$\dot{n}(s) + f(s) = 0$$
$$\dot{m}(s) + \dot{p}(s) \times n(s) + l(s) = 0$$
$$n(s) = R(s)K_{se}(s)(v(s) - v^*(s))$$
$$m(s) = R(s)K_{bt}(s)(u(s) - u^*(s))$$

Where, the applied force distribution per unit of s is given by f the applied moment distribution per unit of s is l. n and m denote the internal force and moment vectors respectively. The kinematic variables in undeformed reference configuration are given by v^* and u^*. K_{se} is the stiffness matrix for shear and extension and K_{bt} is the matrix for bending and torsion. Now combining the rod model with the Cosserat model for extensible strings to describe the tendons we get –

$$f = f_e + f_t$$
$$l = l_e + l_t$$

Where, f_e and l_e are truly external distribution loads and f_t and l_t are distribution loads due to tendon tension. The tendon location in the body coordinate frame is given by –

$$r_i(s) = [x_i(s) \quad y_i(s) \quad 0]^T$$

The parametric space curve defining the tendon path in deformed state is given by –

$$p_i(s) = R(s)r_i(s) + p(s)$$

The distribution loads due to tendon tension is given by –

$$f_t = -\sum_{i=1}^{n} \tau_i \frac{{p_i}'}{||\dot{p}_i||^3} \ddot{p}_i$$

$$l_t = -\sum_{i=1}^{n} \tau_i (Rr_i) \frac{{p_i}'}{||\dot{p}_i||^3} \ddot{p}_i$$

After expressing the force and moment distributions in kinematic variables, we finally get –

$$\begin{bmatrix} \dot{v} \\ \dot{u} \end{bmatrix} = \begin{bmatrix} K_{se} + A & G \\ B & K_{bt} + H \end{bmatrix}^{-1} \begin{bmatrix} d \\ c \end{bmatrix}$$

III. CONTINUUM ROBOT MODEL

For this study we work with robot with spring steel backbone with acrylic spacer disks and a single Kevlar tendon. To create an experimentally accurate model of tendon robot, the parameters are taken from [31]. The rod parameters were calibrated using weight release trial method. The properties of the steel backbone continuum robot are shown in Table. 1.

TABLE I. TENDON ROBOT PARAMETERS

Properties	Variable	Value
Backbone Length	$\boldsymbol{L\ (m)}$	$\mathbf{0.4}$
Backbone Diameter	r (m)	0.0009
Tendon offset	l (m)	0.019
Mass of system	$g\ (\frac{m}{s^2})$	[9.81 0 0]
Direction of gravity	ρ	39317
Density	$E\ (Pa)$	$207e^9$
Elastic Modulus	$G\ (Pa)$	$7.96e^{10}$
Shear Modulus	$L\ (m)$	0.4

End effector position and orientation are computed for various tendon tension representations by solving static equations using the ode45 solver. The routines are written in MATLAB and solved with the "fsolve" using Levenberg Marquardt solver. Following that, the rotation matrix is transformed to quaternions. Quaternions are versatile and don't have parameterization singularities (gimble lock).

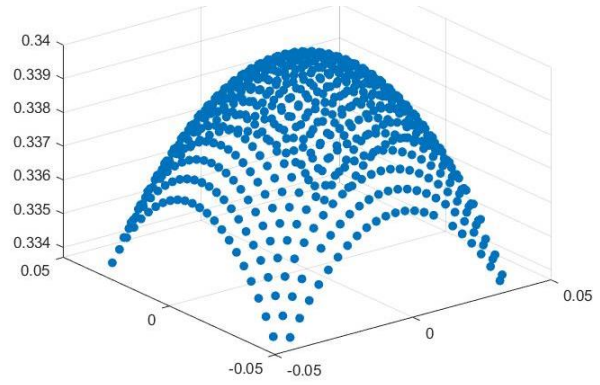


Figure 1. Workspace of the continuum robot

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

Quaternions also provide smooth orientation interpolation and efficient rotation composition. The reachable places of the robot end effector are referred to as the robot workspace. The point cloud in Fig. 1 displays the end effector position, which was mapped to configuration space using a 20000 set of randomized tendon tensions.

IV. CASCADE CORRELATION

In artificial neural networks, during the backpropagation step, the weights in the network are changing at once, which leads to the moving target problem. This leads to slow learning as the hidden units observe a constantly changing environment. Cascade correlation algorithm as shown in Fig. 2 was proposed to solve the problem [32].

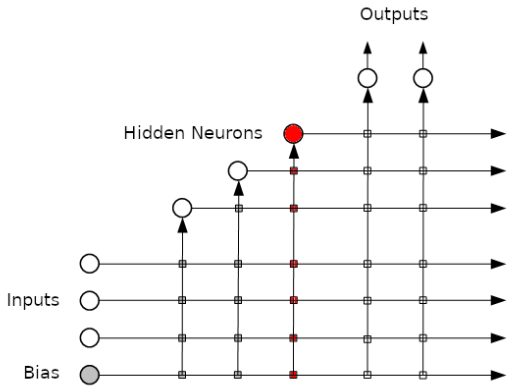


Figure 2. Cascade Network Architecture

Cascade correlation network as shown in Fig starts with the minimal network and then train and add new hidden units automatically. Initially, the input and output are connected with adjustable weights. There is also a bias, which is permanently set to +1. The output of the given configuration is non-linear sum of their weighted inputs.

Following this, hidden units are added one by one. Each hidden unit receives inputs from existing hidden units and origin inputs. Once a unit is added, the input weights are frozen and only the output weights are trained repeatedly. The output weights are trained using the Fahlman's quickprop algorithm. Once the training reaches an asymptote, forward propagation is done one last time to measure the training error. Training stops if we are satisfied with the residual error, if we are not hidden units are added.

To create a new hidden unit, multiple candidate units are evaluated. Each candidate unit is connected to original inputs and existing hidden units. The output of the candidate unit is not connected to the output unit. The weights of the candidate unit are adjusted upon training with the goal to maximise S, S is defined as the sum of sum over all output units o of the magnitude of the correlation (or, more precisely, the covariance) between V, the candidate unit's value, and E_o, the residual output error observed at unit o. S is given by –

$$S = \sum_o \left| \sum_p (V_p - \bar{V})(E_{p,o} - \overline{E_o}) \right|$$

To maximise S, the partial derivate wrt weights is taken and using gradient descent S can be maximised. The candidate unit is chosen whose correlation is the best. The gradient is defined as –

$$\partial S / \partial w_i = \sum_{p,o} \sigma_o (E_{p,o} - \overline{E_o}) f'_p I_{i,p}$$

Where σ_o is the sign of the correlation between the candidate's value and output o, f'_p is the derivative for pattern p of the candidate unit's activation function with respect to the sum of its inputs, and $I_{i,p}$ is the input the candidate unit receives from unit i for pattern p.

Once S stops improving the new candidate is added to the network, the input weights are frozen and the process is continued. As the candidate unit develops a positive correlation with the magnitude of the error, it develops a negative correlation with the weights of the unit, therefore cancelling some of the error. Using a pool of candidates is beneficial as it reduces the chances of installing a useless unit into the network. Since the candidates do not interact with each other and they receive the same input units, training speed can be improved by using parallel machines.

V. RESULTS

To put the model to the test, 20000 random tendon stresses are collected. The network output is used to calculate the end effector location and orientation in quaternions. The dataset is normalized between 0 and 1 to increase model accuracy and speed up the training process. The network has been trained on 97% of the dataset.

Network Performance

Cascade correlation network was trained for 1000 iterations with candidate units selected over 100 iterations. The RMSE error obtained over the training period is shown in Fig. The network was best trained with 10 hidden units and sigmoid function as the activation function. 5 models of the past namely Multi-layer Perceptron, Radial Basis Function, K Nearest Neighbors Regression, Decision Tree, Extreme Learning Machine etc. are modelled to validate the effectiveness and accuracy of our proposed network. The RMSE errors on both training and testing dataset is shown in Table II. Cascade Correlation with sigmoid activation functions is compared against their results and outperforms with significantly lower RMSE values, 1.34 and 1.45, both

TABLE II. RESULTS OF DIFFERENT MODELS

Model	RMSE Training	RMSE Testing	Max Error (mm)
KNNR	1.51	1.87	7.83
RBF	2.83	2.99	6.45
MLP	2.65	2.69	6.25
ELM	1.76	1.84	5.85
Decision Tree	1.17	1.97	11.00
Cascade Correlation	**1.34**	**1.45**	4.13
Cascade Forward backprop	1.68	1.70	5.90

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 11 Issue 04, April-2022

on train and test data. Figure 3 visually compares the RMSE errors of the literature and proposed models, showing improvement.

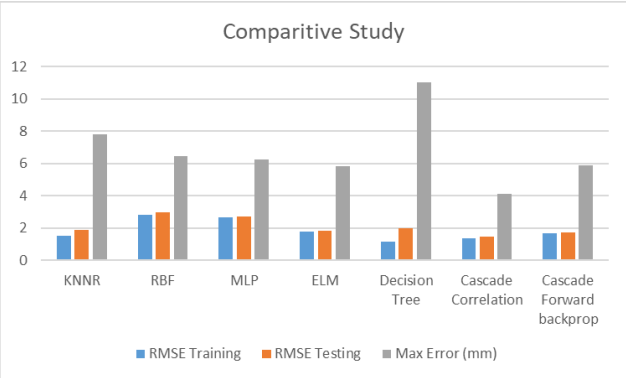


Figure 3. Results of different models compared; Cascade Correlation outperforms other networks

VI. Conclusion

Continuum robot manipulators though complex to control and analyze, offer flexibility and dexterity for a wide array of tasks. In this paper, a static model based on Cosserat model is presented for a single Kevlar tendon. The inverse kinematic of the proposed robot model is obtained through a cascade correlation neural network. Forward kinematic equations are first derived using Cosserat rod model, and datasets containing the robot end-effector in Cartesian space and the tendon tensions are prepared accordingly. The network is then constructed as the IK solver, and the performance is evaluated under different model parameters. From the simulation, a mean squared error of 1.34 can be achieved. The proposed model is then compared to previous literature, validating significant improvement in prediction accuracy.

REFERENCES

[1] Robinson, G., & Davies, J. (1999). Continuum robots - a state of the art. Proceedings 1999 IEEE International Conference on Robotics and Automation

[2] Grzesiak, A., Becker, R., & Verl, A. (2011). The Bionic Handling Assistant: a success story of additive manufacturing. Assembly Automation, 31(4), 329–333.

[3] Hu, H., Wang, P., Zhao, B., Li, M., & Sun, L. (2009). Design of a novel snake-like robotic colonoscope. 2009 IEEE International Conference on Robotics and Biomimetics (ROBIO).

[4] Laschi, C., Mazzolai, B., Mattoli, V., Cianchetti, M., & Dario, P. (2009). Design of a biomimetic robotic octopus arm. Bioinspiration & Biomimetics, 4(1), 015006.

[5] Camarillo, D. B., Krummel, T. M., & Salisbury, J. (2004). Robotic technology in surgery: Past, present, and future. The American Journal of Surgery, 188(4), 2–15.

[6] Tsukagoshi, H., Kitagawa, A., & Segawa, M. (2001). Active Hose: an artificial elephant's nose with maneuverability for rescue operation. Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation

[7] Godage, I. S., Branson, D. T., Guglielmino, E., & Caldwell, D. G. (2012). Pneumatic muscle actuated continuum arms: Modelling and experimental assessment. 2012 IEEE International Conference on Robotics and Automation.

[8] Webster, R. J., Romano, J. M., & Cowan, N. J. (2009). Mechanics of Precurved-Tube Continuum Robots. IEEE Transactions on Robotics, 25(1), 67–78.

[9] Seok, S., Onal, C. D., Cho, K. J., Wood, R. J., Rus, D., & Kim, S. (2013). Meshworm: A Peristaltic Soft Robot with Antagonistic Nickel Titanium Coil Actuators. IEEE/ASME Transactions on Mechatronics, 18(5), 1485–1497.

[10] Camarillo, D., Carlson, C., & Salisbury, J. (2009). Configuration Tracking for Continuum Manipulators with Coupled Tendon Drive. IEEE Transactions on Robotics, 25(4), 798–808.

[11] Hannan, M. W., & Walker, I. D. (2001). Analysis and experiments with an elephant's trunk robot. Advanced Robotics, 15(8), 847–858.

[12] Dermatas, E., Nearchou, A., & Aspragathost, N. (1996). Error-back-propagation solution to the inverse kinematic problem of redundant manipulators. Robotics and Computer-Integrated Manufacturing, 12(4), 303–310.

[13] Rone, W. S., & Ben-Tzvi, P. (2014). Continuum Robot Dynamics Utilizing the Principle of Virtual Power. IEEE Transactions on Robotics, 30(1), 275–287.

[14] Jones, B., & Walker, I. (2006). Kinematics for multisection continuum robots. IEEE Transactions on Robotics, 22(1), 43–55.

[15] He, B., Wang, Z., Li, Q., Xie, H., & Shen, R. (2013). An Analytic Method for the Kinematics and Dynamics of a Multiple-Backbone Continuum Robot. International Journal of Advanced Robotic Systems, 10(1), 84.

[16] Li, M., Kang, R., Geng, S., & Guglielmino, E. (2017). Design and control of a tendon-driven continuum robot. Transactions of the Institute of Measurement and Control, 40(11), 3263–3272.

[17] Jones, B. A., Gray, R. L., & Turlapati, K. (2009). Three dimensional statics for continuum robotics. 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems.

[18] Rucker, D. C., & Webster III, R. J. (2011). Statics and Dynamics of Continuum Robots with General Tendon Routing and External Loading. IEEE Transactions on Robotics, 27(6), 1033–1044.

[19] Song, S., Li, Z., Yu, H., & Ren, H. (2015). Shape reconstruction for wire-driven flexible robots based on Bézier curve and electromagnetic positioning. Mechatronics, 29, 28–35.

[20] Ding, J., Goldman, R. E., Xu, K., Allen, P. K., Fowler, D. L., & Simaan, N. (2013). Design and Coordination Kinematics of an Insertable Robotic Effectors Platform for Single-Port Access Surgery. IEEE/ASME Transactions on Mechatronics, 18(5), 1612–1624.

[21] Kuntz, A., Sethi, A., Webster, R. J., & Alterovitz, R. (2020). Learning the Complete Shape of Concentric Tube Robots. IEEE Transactions on Medical Robotics and Bionics, 2(2), 140–147.

[22] Loutfi, I. M., Boutchouang, A. B., Melingui, A., Lakhal, O., Motto, F. B., & Merzouki, R. (2020). Learning-Based Approaches for Forward Kinematic Modeling of Continuum Manipulators. IFAC-PapersOnLine, 53(2), 9899–9904.

[23] Shahabi, E., & Kuo, C. H. (2018). Solving Inverse Kinematics of a Planar Dual-Backbone Continuum Robot Using Neural Network. EuCoMeS 2018, 355–361.

[24] Wu, G., Shi, G., & Shi, Y. (2017). Modeling and analysis of a parallel continuum robot using artificial neural network. 2017 IEEE International Conference on Mechatronics (ICM).

[25] Lai, J., Huang, K., & Chu, H. K. (2019). A Learning-based Inverse Kinematics Solver for a Multi-Segment Continuum Robot in Robot-Independent Mapping. 2019 IEEE International Conference on Robotics and Biomimetics (ROBIO).

[26] Shen, W., Yang, G., Zheng, T., Wang, Y., Yang, K., & Fang, Z. (2020). An Accuracy Enhancement Method for a Cable-Driven Continuum Robot with a Flexible Backbone. IEEE Access, 8, 37474–37481.

[27] Xu, W., Chen, J., Lau, H. Y., & Ren, H. (2016). Data-driven methods towards learning the highly nonlinear inverse kinematics of tendon-driven surgical manipulators. The International Journal of Medical Robotics and Computer Assisted Surgery, 13(3)

[28] Thuruthel, T. G., Falotico, E., Cianchetti, M., & Laschi, C. (2016). Learning Global Inverse Kinematics Solutions for a Continuum Robot. ROMANSY 21 - Robot Design, Dynamics and Control, 47–54.

[29] Rucker, D. C., & Webster III, R. J. (2011). Statics and Dynamics of Continuum Robots With General Tendon Routing and External Loading. *IEEE Transactions on Robotics*, *27*(6), 1033–1044.

[30] Jones, B. A., Gray, R. L., & Turlapati, K. (2009). Three dimensional statics for continuum robotics. *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*.

[31] Till, J., Aloi, V., & Rucker, C. (2019). Real-time dynamics of soft and continuum robots based on Cosserat rod models. The International Journal of Robotics Research, 38(6), 723–746.

[32] Liang, H. (1998). Improvement of cascade correlation learning. *Information Sciences*, *112*(1–4), 1–6.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

View publication stats

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360610606

Multi –Objective Parametric Optimization of Milling on AMMC using Supervised Machine Learning Techniques

Article *in* Journal of Critical Reviews · May 2022

DOI: 10.15680/IJIRSET.2022.1104126

CITATIONS
0

READS
4

4 authors, including:



A. Ospanova
Al-Farabi Kazakh National University
248 PUBLICATIONS **593** CITATIONS

SEE PROFILE

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

IJIRSET APRIL View project

IJIRSET APRIL View project

All content following this page was uploaded by A. Ospanova on 15 May 2022.

The user has requested enhancement of the downloaded file.

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

Multi – Objective Parametric Optimization of Milling on AMMC using Supervised Machine Learning Techniques

Dr. A.K. Madan[1], Pratyush Bhatt[2], Pranav Taneja[3]

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

Department of Mechanical Engineering, Delhi Technological University, Delhi, India[2,3]

ABSTRACT: Aluminium Metal Matrix Composites (AMMC) have recently gained importance inproduction technology.The presence of hard reinforcements in the composites makes the machining process (milling, turning and drilling) quite difficult.Experimentation determination of optimum parameters for machining AMMC is expensive and time intensive.This can be substituted by using machine learning techniques.This study aims to optimize the input parameters such as depth of cut, feed and cutting speed in the milling for differentmass fraction and particle size of SIC. This is achieved by predicting the output parameters by using supervised machine learning techniques - LASSO Regression, KNN Regressionand Artificial Neural Network Regression.The results of the various models are compared using metrics like MSE and R^2.The sustainable impacts of incorporating artificial intelligence with manufacturing techniques(Industry 4.0) are thus demonstrated.

KEYWORDS: Industry 4.0, Sustainable manufacturing, CNC Milling, Supervised Machine Learning, AMMC.

I. INTRODUCTION

Aluminium Metal Matrix Composites (AMMCs) are a class of materials which find extensive applications where light-weight, high stiffness and moderate strength are required properties [1]. Metal matrix composites are metals reinforced with other metals, ceramics or organic compounds. Reinforcements are dispersed in the matrix and are majorly included to improve or enhance the properties of base metal like strength, stiffness, conductivity, etc. Aluminium as base metal has attracted the most attention in such composites. The reinforcements must be stable and non-reactive in the working temperature. Silicon Carbide (SiC) and Aluminium Oxide (Al_2O_3) are the most commonly used reinforcements; SiC increases the tensile strength, hardness, density and wear resistance of aluminium alloys while Al_2O_3 imposes excellent compressive strength and wear resistance. Boron Carbide's (B_4C) addition also enhances the hardness but doesn't improve the wear resistance [2]. Thus, with a variety of reinforced materials and flexibility in their primary processing, AMMCs offer great potential for the development of composites with the desired properties for certain applications [3]. In the last few years, AMMCs have largely replaced the conventional ferrous materials in automobile, marine, and aerospace industries owing to its higher wear resistance and strength to weight ratio [4].

Milling is one of the most extensively used machining process which may be employed in at least one stage of fabrication in the manufacturing industries. CNC Milling machines are most commonly used as they are versatile, flexible and allow the manufacturing of products in shorter duration of time at reasonable costs along with great surface finish [5]. It is also capable of producing complex geometric surfaces with reasonably high accuracy and surface finish.Milling process removesmaterial by cutting repeatedly into small bits.Milling cutter is a rotary tool with multiple cutting points and it is usually moved perpendicular to the axis of workpiece so that cutting occurs on the circumference only [6]. The various input or response parameters involved in CNC Milling are explained below:

Particle size: - Particle size is defined as the dimensions of reinforcements in aluminium metal matrix composites. It is measured in µm. SiC particles of 10, 20, 40 µm were reinforced [7].

Mass Fraction: - Mass Fraction is defined as the ratio of weight percentage of the reinforcements used to that of the metal matrix composite.

Feed Rate: -It is the speed with which the cutting tool is fedagainst the workpiece.It is measured in mm/min.

Depth of Cut: - Depth of cut also known as Axial or Radial Depth of cut is the total thickness of the cut when viewing the cutting tool from above. It can never exceed the cutter's diameter and is measured in mm.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

Cutting Velocity: - Cutting velocity is the tangential velocity at the periphery of rotating cutter or workpiece. It is measured in rpm.

The output variables obtained in the CNC milling process are:

Surface Roughness (Ra): - Measure of irregularities on machined surfaces or the measure of vertical aberrations along a flat surface. It is measured in μm using a surface roughness comparator or profilometer.

Material Removal Rate (MRR): - Material removed per unit time. It is measured in mm^3/min.

$$MRR = kerf\ width\ (k) * thickness\ of\ workpiece\ (t) * cutting\ velocity\ (V_c)$$

Experimental determination of the optimum input parameters for desired Ra and MRR in CNC milling is an effective and accurate method. But due to the slow rate of MRR, a lot of time is consumed in the process, especially for large sized workpieces. Also, excessive power is consumed in the milling operation. Moreover, a large capital is required for the experimentation. Substitution of experimental analysis by supervised machine learning algorithms for the prediction of Ra and MRR resolves these issues. In Supervised Machine Learning, computational or mathematical algorithms are trained using "labelled" training data. Labelled data comprises a collection of input variables along with their corresponding output variables. This enables the algorithms to predict the output for unseen data having input and output fields similar to the training data. Therefore, the aim of supervised machine learning algorithm is to generate a mapping function that can accurately match the input variables to the output variables. The performance of machine learning models isdetermined using Evaluation Metrics. It is necessary to use multiple evaluation metrics to evaluate the model. This is because a model may perform well on one evaluation metric, but may perform poorly on another. Evaluation metrics are used for critical analysisto ensure that the model functions optimally. So, we have used MSE metric and R^2 score to check the validity of the model. Mean Squared Error is the mean of square of errors produced. It is calculated by

$$\text{MSE} = \sum_{i=1}^{n} (y_i - \hat{y})^2$$

The lower the MSE score, the better is the model. R^2 score is a statistical measure that represents the proportion of the variance for a dependent variable that's explained by an independent variable or variables in a regression model. It is given by

$$R^2 = 1 - \frac{RSS}{TSS}$$

RSS standsfor residual sum of squares andTSS stands for total sum of squares. The higher the R^2 score, the better is the model.

The objective of the study is to map the response parameters – Particle Size, Mass Fraction, Feed Rate, Depth of Cut and Cutting Velocity to Surface Roughness and MRR using several supervised machine learning models - LASSO Regression, KNN regression and Neural Network regression. Thus, optimization of multiple input parameters (Multi - Parametric optimization) for desired output parameters - Surface Roughness and MRR is performed simultaneously (Multi - Objective optimization) for milling of AMMC. The performance of ML models is evaluated using the metric - Mean Squared Error (MSE) and R^2 score, and the technique which most accurately models the milling process, for such a small dataset is identified.

II. RELATED WORK

In recent years, many researches have been performed to optimize response parameters of CNC milling process for different materials either using analytical Taguchi's method or machine learning techniques. K.G. Prasad utilized Data Envelopment Analysis (DEA) based Taguchi method in order to optimize CNC End Milling Process Parameters. V. Mundada et.al performed the optimization of milling operations using Artificial Neural Network (ANN) and Simulated Annealing Algorithm (SAA) and Response Surface Methodology (RSM) was used to formulate its objective function [8]. P. Saini et.al used Taguchi's Grey Relational Analysis so as to optimize end milling parameters for rough and finish machining of Al-4032 with 3% SiC metal matrix composite [9]. E. Mounayri et.al provided a detailed study on multi parametric ANN model for flat milling of aluminium grade 7075T6 [10]. Daniel et. al optimised the control parameters in CNC milling of AMMC using ANN and used Taguchi -grey relational analysis for optimization [7].

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

III. METHODOLOGY

3.1 Dataset Description

The dataset was obtained from milling experiments onAluminumhybrid MMC from literature [7]. The dataset consists of 27 runs or experiments with 3 levels of 5 input parameters –Cutting Speed, Feed rate, Depth of cut, mass fraction and particle size of SiC andthe corresponding 2 output parameters – R_a and MRR obtained from the experiments. The response parameters along with their levels are:

Levels	SIC (%)	SIC Size (micrometres)	Cutting Speed (Rpm)	Feed Rate (mm/min^{-1})	Depth of Cut (mm)
1	5	10	500	200	0.5
2	10	20	1000	400	1
3	15	40	1500	600	1.5

Table 2.1. Four levels of response parameters

3.2 Supervised Machine Learning Models

3.2.1 LASSO Regression

Least Absolute Shrinkage and Selection operator (LASSO)regression is form linear regression that uses shrinkage. Shrinkage operation promotesdata values to get shrunktowards a central point such as the mean. The lasso procedure thus encourages simple, sparse models (i.e. models with fewer parameters). Lasso regresion achieves this by perfroming regularization, in which a penalty term is added to residual sum of squares term [11]. For LASSO, absolute magnitude of coefficents is taken as the penalty. This is known as theL1 regularization technique [12].The objective of the algorithm is to minimize [13]:

$$\sum_{i=1}^{n}(y_i - \sum_{j} x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$

Where λ is a tuning parameter that regulates the effect of L1 penalty term [14]. When some of the βs are minimized to zero, the regression model becomes computationally cheap and easier to interpret. Thus, LASSO regression is mostly employed in models having large multicollinearity.

3.2.2 K Nearest Neighbors Regression

The KNN algorithm uses “feature similarity” in order to determine new data points. The value of a new point is evaluated based on the resemblance to the points in training set. The average of the values is taken as final prediction. KNN regression can be explained as a non-parametric method which intuitivelyformulates a relation between independent variables and the continuous outcomes by averaging observations belonging to the same neighborhood [15]. Size of neighborhood is chosen using cross-validation in which sizeproducing the least mean-squared error is selected. Our KNN model uses Euclidean distance metric which is calculated as [16]:

$$\sqrt[2]{\sum_{i=1}^{k}(x_i - y_i)^2}$$

Where x is a point in the training set, y is a point in test set and K is the number of neighbors. The method is unfeasible for data having large dimensions ornumerous independent variables. For a low k-value, overfitting occurs on training set leading to poor results on test set [17]. On the other hand, for a high k-value, model produces poor

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

results on testing as well as training dataset. Thus, the error curve is minimum only for a specific k-value. It is evaluated as 4 for the milling model.

3.2.3 Artificial Neural Network Regression

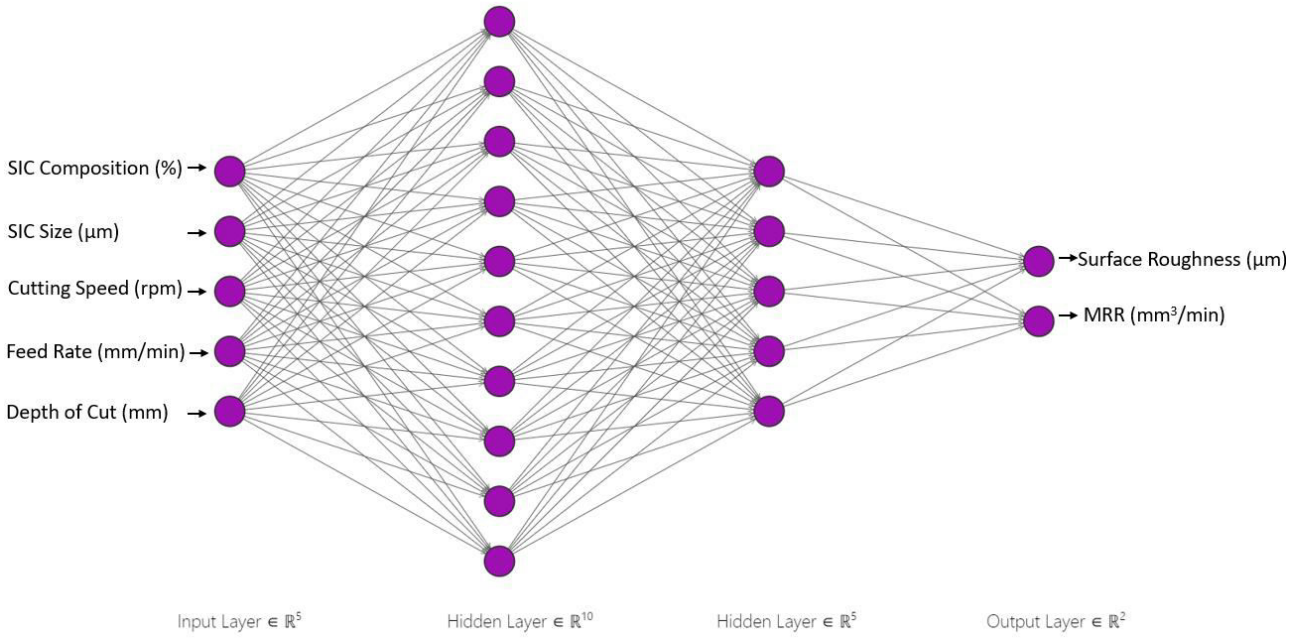


Figure. 1. Artificial Neural Network Structure for modelling milling

Artificial Neural Network (ANN) is a set of computational functions which helps to mapthe input data to the output. Neural networks consist of neurons arranged in sequential layers having weights and biases assigned to them. Weights and biases are learnable parameters that are multiplied and added respectively with the inputs received to the neuron from the previous layer. These weights are updatedas the learning progresses [18]. In the milling model there are four layers, the Input layer consisting of 5 neurons (for the five input parameters), the Output layer providing the required outputs consisting of 2 neurons (for R_a and MRR), and 2 Hidden layers for processing, consisting of 10 and 5 neurons. The output generated from each neuron is passed to an activation function (ReLU) which then passes the obtained output to next layer.

$$Layer\ Output = f(\sum_{i=1}^{n} w_i x_i + b_i)$$

Where x is the input to neuron, w and b are the weights and biases, n is the number of inputs to this neuron and f () is the activation function [19]. On each iteration during training of the network, Gradient based optimizers(Adam) adjust the weights to improve the accuracy by minimizing the error obtained by loss function (Mean squared error). The training process is repeated for 2000 epochs after whichMSE does not reduce further.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

IV. EXPERIMENTAL RESULTS

4.1 LASSO Regression

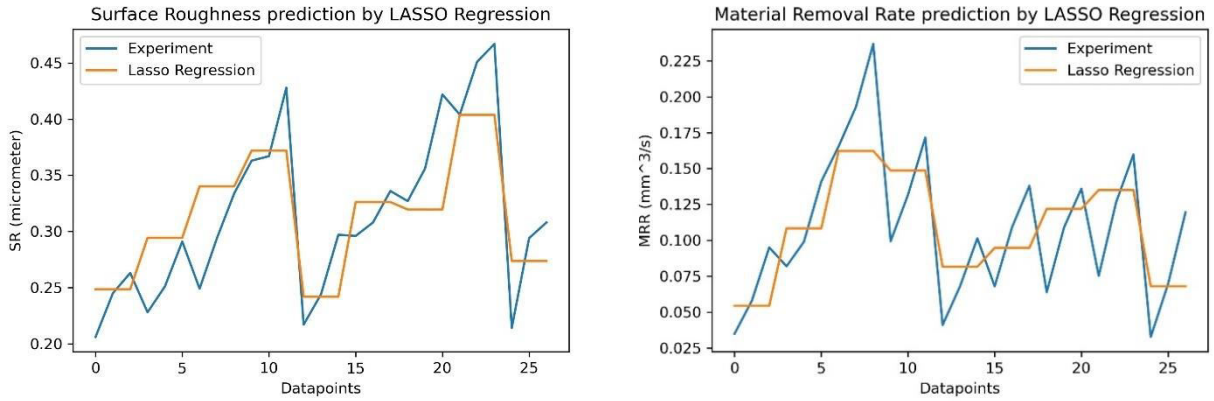


Figure. 2. Actual (experiments) and predicted values (LASSO Regression) for R_a (left) and MRR (right)

The MSE values for R_a and MRR produced by LASSO regression are 0.0018and 0.0011 respectively. The R^2 score for R_a and MRR are 0.65 and 0.52 respectively. Though the error is highly reduced (Fig 2), the predicted values do not overlap with the extreme experimental values present in dataset (Fig 2). This happened because LASSO regression objective is to minimize the residual sums of square loss and thus ends up predictingvalues near to the mean of the dataset. This way a low MSE value is produced but the prediction is not correct for data points in which the values deviate significantly from the mean.

4.2 K – Nearest Neighbor Regression

The MSE values for R_a and MRR produced by KNN regression are 0.0021 and 0.0008 respectively. The R^2 score for R_a and MRR are 0.61 and 0.65 respectively. The error is quite high for most data points and the predicted values do not overlap with experimental values as well (Fig 3). This occurred because KNN regression requires large amounts of good quality data. Distribution of all datapoints very closely deteriorates the results of KNN.

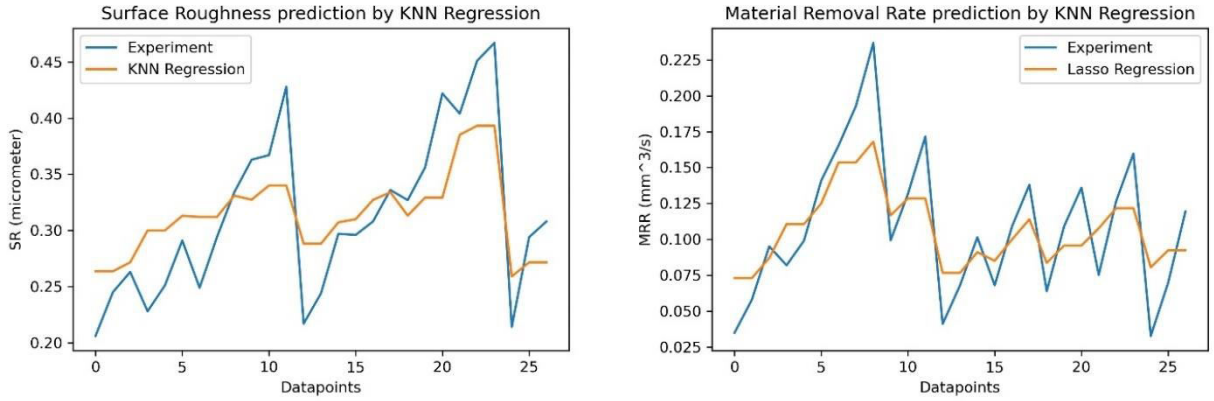


Figure. 3. Actual (experiments) and predicted values (KNN) for R_a (left) and MRR (right)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

4.3 Artificial Neural Network Regression

The MSE values for R_a and MRR produced by ANN are 0.0015 and 0.0005 respectively. The R^2 score for R_a and MRR are 0.71 and 0.76 respectively. The error values are minimized and the predicted values overlap with the experimental values as well (Fig 4). Thus, ANN accurately models the milling process, even for such a small dataset. ANN is capable of modeling complex nonlinear functions because of different weights and biases associated with the neurons and the activation functions of the layers. After the neural network is trained for a sufficient number of epochs, the optimizer is able to set the weights so as to precisely map any continuous function, which has been proved using the universal approximation theorem [20]. Accuracy of the model will further increase with increase in size of dataset and possible overfitting will also be avoided.

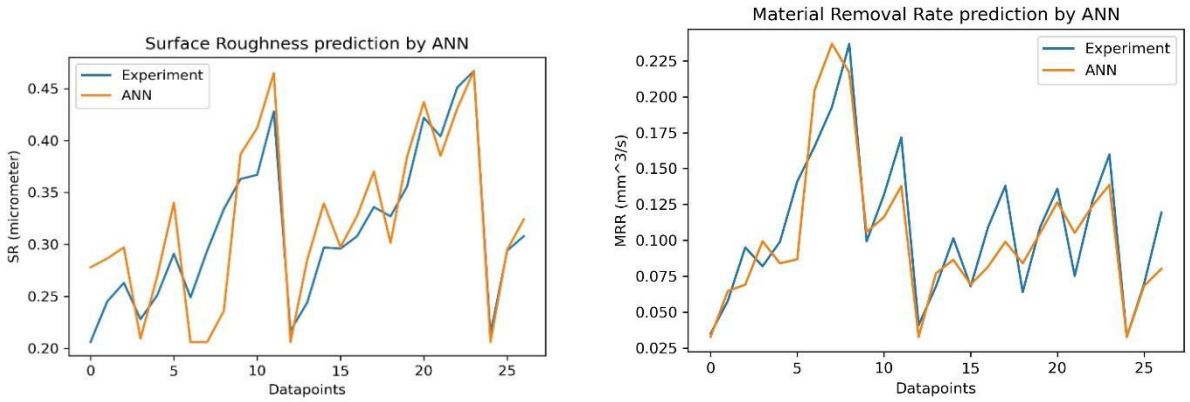


Figure. 4. Actual (experiments) and predicted values (ANN) for R_a (left) and MRR (right)

V. CONCLUSION

The mean squared error for the predicted surface roughness and MRR values are 0.0018, 0.0011 for LASSO regression, 0.0021, 0.0008 for KNN and 0.0015, 0.0005 for ANN respectively.The R^2 Score for the predicted surface roughness and MRR values are 0.65, 0.52 for LASSO regression, 0.61, 0.65 for KNN and 0.71, 0.76for ANN respectively. Among all the supervised machine learning algorithms utilized in the study, the Artificial Neural Network model predicted the surface roughness as well as MRR with highest accuracy or overlap, least MSE and produced the highest R^2 Score. Although training of neural networks requires a greater computation time and power than other algorithms, the difference is insignificant for smaller datasets. Therefore, it can be concluded that artificial neural networks are capable of accurately modelling the CNC Milling process and effectively replaces the experimental analysis required for the optimum parameter determination. The study thus highlights the sustainable impacts of incorporating artificial intelligence with manufacturing techniques, which not only cut-down the time and costs required for the experimentation but also produces parts with high quality surface finish and dimensional accuracy.

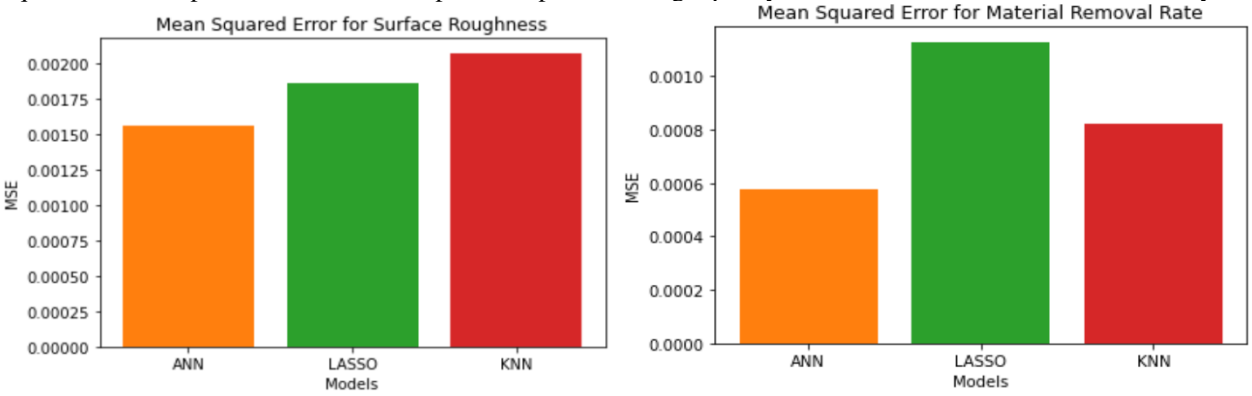


Figure. 5. MSE between actual (experiments) and predicted values (ML models) for R_a (left) and MRR (right)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 8.118|

||Volume 11, Issue 4, April 2022||

| DOI:10.15680/IJIRSET.2022.1104126 |

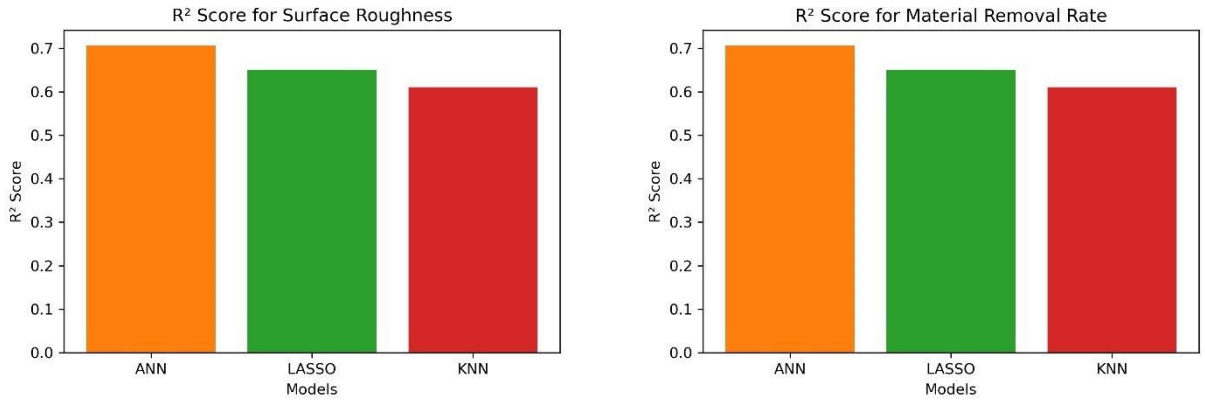


Figure. 6. R^2 between actual (experiments) and predicted values (ML models) for Ra (left) and MRR (right)

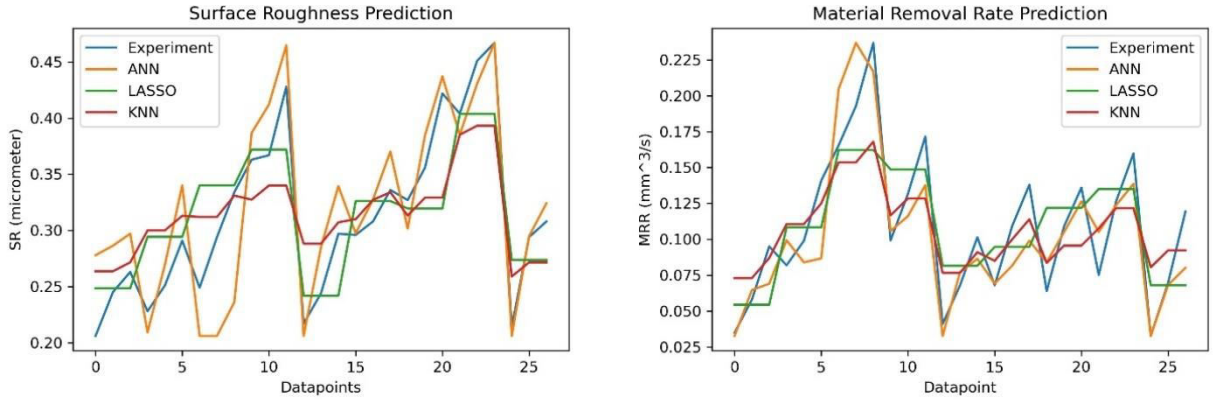


Figure. 4.2. Actual (experiments) and predicted values (ML models) for R_a (left) and MRR (right)

REFERENCE:

1. Кобилов, А. У., & Джурабоев, А. М. (2020). Технологии компьютерного дистанционного обучения. Academic Research in Educational Sciences, 1 (3), 287-293.
2. Кобилов А. У., Ольховская, И. В. (2021). Инновации–форма управленческой деятельности государства и организаций. Экономика и бизнес теория и практика, 1(1), 173-178.
3. Зияев, Т. М., Ўринович, Қ. А., Ёқуб, Д. Т. (2019). Қулай инвестицион муҳит ва фаол инвестицион сиёсат–барқарор иқтисодий ўсиш омили. Экономика и финансы (Узбекистан), 2(10), 555-559.
4. Kobilov, A. U., Khashimova, D. P., Mannanova, S. G., Abdulakhatov, M. M. O. (2022). Modern Content and Concept of Digital Economy. International Journal of Multicultural and Multireligious Understanding, 9(2), 375-378.
5. Kobilov, A. (2021). The concept of digital economy, its features and prospects. Архив научных исследований, 2(1), 42-48.
6. Akhmedov, B. A. (2022). Analysis of the Reliability of the Test form of Knowledge Control in Cluster Education. Psychology and Education, *59*(2), 403-418.
7. Муратов, К. (2021). Анализ сложноподчиненных предложений при изучении синтаксиса. Academic research in educational sciences, 2(4), 1371-1378.
8. Муратов, К. К. (2020). Методика преподавания темы «словообразование» учащимся групп с узбекским языком обучения. Science and Education, 1(Special Issue 2).
9. Курбанович, М. К. (2020). Инновационная деятельность как обязательный компонент личной педагогической системы преподавателя. Суз санъати,(3), 219-225.
10. Муратов, К. (2021). Основы традиций и инноваций в образовательном процессе. Экономика и социум, 11(90), 974-978.
11. Dilrabo Saydakhmetova. (2020). Teaching Russian as a Foreign Language Through Stories. European Journal of Research and Reflection in Educational Sciences, 8(5), 75-78
12. Сайдахметова, Д. Х. (2020). Международные оценочные программы (PISA, TIMSS, PIRLS) и подготовка учеников к ним. Science and Education, 1(Special Issue 2), 112-120.
13. Сайдахметова, Х. Д. (2021). Качества обучения в системе вузовского образования. Academic research in educational sciences, 2(1), 1148-1152.
14. Khasanovna, S. D. (2021). Education of feelings with artistic words. International Engineering Journal for Research & Development, 6(4), 234-238.
15. Сайдахметова, Д. Х. (2020). Цифровая технология. Academic Research in Educational Sciences, 1(4), 85-90.
16. Сайдахметова, Д. Х. (2021). Роль заимствований в межкультурной коммуникации. Academic research in educational sciences, 2(4), 864-868.
17. Сайдахметова, Д. Х. (2022). Место и роль перевода в современной жизни. Ижтимоий фанларда инновация онлайн илмий журнали, 2(4), 96-99.
18. Saidakhmetova, D. K. (2022). International Organization (IEA). International Journal of Innovative Research in Science Engineering and Technology, 11(2), 1599-1602.
19. Saidakhmetova, D. K. (2022). The Concept of the Development of Modern Translation the Place and Role of Translation in Modern Life. International Journal Of Multidisciplinary Research In Science Engineering and Technology, 5(3), 455-456.
20. Khasanovna, S. D. (2021). Semantic features of Russian borrowings in the modern uzbek literary language. Emergent Journal of Educational Discoveries and Lifelong Learning (EJEDL), 2(05), 35-38.
21. Сайдахметова, Д. Х. (2021). Роль перевода в современной жизни. Экономика и социум, 3-2, 298-300.
22. Сайдахметова, Д. Х. (2020). Концепция развития современного переводоведения, место и роль перевода в современной жизни. НАЗАРИЯ АМАЛИЁТ ИННОВАЦИЯ, 1(27), 90-93.
23. Сайдахметова, Д. Х. (2020). Художественный текст как единица обучения чтению. Преподавание языка и литературы, 4(8), 80-87.
24. Сайдахметова, Д. Х. (2020). Повышение эффективности качества высшего образования: проблемы и решения. ИННОВАЦИЯ ВА ИСТИҚБОЛЛАР, 3(9), 29-31.
25. Саиджалилова, Д. Д., Уразова, Ш. Б., Мирзаева, Д. Б. (2021). Анализ генетического полиморфизма генов у женщин с невынашиванием беременности при экстракорпоральном оплодотворении. Клиник ва назарий тиббиёт, 1(6), 104-105.
26. Акпербекова, И. С., Аюпова, Ф. М., Саиджалилова, Д. Д., Саттарова, К. А., Миродилова, Ф. Б. (2020). Роль и значение маточных естественных киллеров в течении здоровой беременности (обзор литературы). Problemy Reproduktsii, 26(6), 15-25.
27. Бобоев, К. Т., Саиджалилова, Д. Д., Ходжаева, Д. Н., Мирзаева, Д. Б. (2020). Роль полиморфизма тромбофилических генов гемостаза в невынашивании беременности при экстракорпоральном оплодотворении (эко). Новости дерматовенерологии и репродуктивного здоровья, 1(2), 61-63.
28. Ayupova, F. M., Saidjalilova, D. D., Muminova, Z. A., Hodjaeva, D. N. (2020). The nature of distribution of fetal extracellular dna during physiological pregnancy. Central Asian Journal of Medicine, 2(1), 20-28.
29. Ayupova, F., Mirzaeva, D., Saidjalilova, D. (2019). EP1062 Hyperplastic processes in hormone-dependent organs in women after IVF. International Journal of Gynecologic Cancer, 29(Suppl 4), 559-560.
30. Ayupova, F., Mirzaeva, D., Saidjalilova, D. (2019). EP664 The effectiveness of the procedure of IVF in patients with uterine myoma. International Journal of Gynecologic Cancer, 29(Suppl 4), 385-386.
31. Рахимов, Р. А., Муминова, З. А., Саиджалилова, Д. Д. (2019). Спектр респираторных вирусов, обусловивших тяжелое течение острой респираторной инфекции у беременных. Журнал теоретической и клинической медицины, 1(1), 113-117.
32. Саиджалилова, Д. Д., Мирзаева, Д. Б. (2019). Изучение роли генетического исследования системы гемостаза в эффективности экстракорпорального оплодотворения. Медицинские новости, 7(298), 60-62.
33. Ayupova, F. M., Saidjalilova, D. D., Muminova, Z. A., Hodjaeva, D. N. (2019). The Interdependence of Fetal and Maternal Extracellular DNA in Pregnant Women with Fetal Malformations. Central Asian Journal of Medicine, 2019(3), 45-50.

View publication stats

Open Access Article

MULTI-FEATURE AWARE POSE AND GEOMETRY BASED FACIAL EXPRESSION RECOGNITION USING DEEP LEARNING

Ankit Sharma
Department of Computer Science and Engineering, Delhi Technological University,
ankitsharma_2k20cse05@dtu.ac.in

Dr. Manoj Sethi
Department of Computer Science and Engineering, Delhi Technological University,
manojsethi@dce.ac.in

1. Abstract

Facial expressions are the primary way to express intentions and emotions. In real life, computers can understand the emotions of humans by analysing their facial expressions. Facial Expression Recognition (FER) plays an important role in human-computer interaction necessity and medical field [1]. In the past, the facial features are extracted manually for recognizing the expressions. In the present, it is an important hotspot in computer vision, Internet of Things, and artificial intelligence fields. Certain processes are involved to recognize the facial expression in an efficient manner such as,

- Preprocessing
- Segmentation
- Feature Extraction
- Classification

Various algorithms are designed manually for feature extraction and other feature extraction algorithms such as Local Binary Pattern (LBP), Gabor wavelet, Histogram of Oriented Gradient (HOG) [2],[3]. Various challenges are involved in the FER to recognize accurate expressions of facial images. For robust classification of facial expression, consideration of illumination and pose of the facial image is important. The poses and facial identity learning are essential to get accurate results. Several existing works faced challenges regarding identity, pose variation, and inter-subject variation. For estimating the pose of the facial images existing methods used hand-crafted features. For detecting the pose of the facial images, pose normalization is performed by considering the angle in the existing works [3]. Previous works also considering pixel-based normalization to increase the accuracy for recognizing the facial expression [2]. On the other hand, general illumination effect of the images also affects the accuracy, by contrast, occlusion, etc. Segmentation is one of the major processes to partition the facial images to extract the features. Various existing methods used different techniques to segment the facial images such as bounding box-based segmentation, region-based segmentation, cluster-based segmentation, etc., by using various algorithms such as Discrete Cosine Transform (DCT), K-means Clustering algorithm, etc. For extracting the features after segmenting the facial images various

Received: January 24, 2022 / Revised: February 22, 2022 / Accepted: March 27, 2022 / Published: April 30, 2022

About the authors：Ankit Sharma

Email: ankitsharma_2k20cse05@dtu.ac.in

Machine learning (ML) algorithms such as Support Vector Machines (SVM), Naive Bayes, K-Nearest Neighbours (KNN), and Deep Learning (DL) algorithms such as Convolutional Neural network (CNN), Generative Adversarial Network (GAN), Long-Short Term Memory Networks (LSTM) [2], [3], [4]. After extracting the features from the facial images, classification is performed to identify the facial expressions (happy, sad, anger etc.). Various classifiers are used in the previous works for classification processes such as VGG-16, VGG-19, ResNet, etc [2], [4]. These classifiers get the extracted feature as input and classify the expressions. For training and testing purposes various datasets are used but, mostly FER-2013, CK+, and JAFFE datasets are used [1], [2], [3].

Keywords: *Facial Expression Recognition, Deep Learning, Capsule Neural Network, Fuzzy C-means Clustering, Segmentation, Facial Landmark Detection*

1. **摘要**

面部表情是表达意图和情绪的主要方式。在现实生活中，计算机可以通过分析人类的面部表情来理解人类的情绪。面部表情识别（FER）**在人机交互的必要性和医学**领域发挥着重要作用[1]。过去，面部特征是人工提取的，用于识别表情。目前，它是计算机视觉、物联网、人工智能领域的重要热点。涉及以有效方式识别面部表情的某些过程，例如，

• 预处理

• 细分

• **特征提取**

• **分**类

为特征提取和其他特征提取算法手动设计了各种算法，例如局部二值模式 (LBP)、Gabor **小波**、**定向梯度直方**图 (HOG) [2]、[3]。 FER **涉及各种挑**战，以识别面部图像的准确表达。对于面部表情的鲁棒分类，考虑面部图像的照明和姿势很重要。姿势和面部识别学习对于获得准确的结果至关重要。一些现有的作品在身份、姿势变化和主体间变化方面面临挑战。为了估计面部图像的姿势，现有方法使用了手工制作的特征。为了检测面部图像的姿势，姿势归一化是通过考虑现有作品中的角度来执行的[3]。**以前的工作**还考虑了基于像素的归一化以提高识别面部表情的准确性 [2]。**另一方面**，图像的一般光**照效果也会影响精度**，**如**对比度、遮挡等。分割是对人脸图像进行分割以提取特征的主要过程之一。现有的各种方法使用不同的技术对人脸图像进行分割，如基于边界框的分割、基于区域的分割、基于聚类的分割等，通过使用各种算法，如离散余弦变换（DCT）、K-means **聚**类算法等用于在分割面部图像后提取特征的各种机器学习 (ML) **算法**，**如支持向量机** (SVM)、**朴素**贝叶斯、K-**最近**邻 (KNN) **和深度学**习 (DL) **算法**，**如卷**积神经网络 (CNN)、**生成**对抗网络 (GAN)、长短期记忆网络 (LSTM) [2]、[3]、[4]。**从面部**图像中提取特征后，进行分类以识别面部表情（快乐、悲伤、愤怒等）。以前的工作中使用了各种分类器来进行分类过程，例如 VGG-16、VGG-19、ResNet **等** [2]、[4]。这些分类器将提取的特征作为输入并对表达式进行分类。出于训练和测试目的，使用了各种数据集，但主要使用 FER-2013、CK+ **和** JAFFE **数据集** [1]、[2]、[3]。

关键词：面部表情识别、深度学习、胶囊神经网络、模糊 C **均**值聚类、分割、面部标志检测

2. Research Aim and Objective

The aim of this research is to identify the facial expression from the facial images by using Deep learning technique. In addition, this research identifies the problems of considering adequate features, high false-positive rate, less accuracy, and so on.

The main objective of this research is to recognize the facial expressions from the facial images with low false positive rate and high accuracy. The remaining research objectives are described as follows,

- To increase the facial image quality preprocessing is initialized by performing normalization to estimate the pose and angle of the facial image and also reduce the illumination effects to increase the accuracy of the facial expression.
- To extract the features efficiently, cluster-based segmentation is performed to estimate the facial objects such as eyebrows, eyes, nose, and mouth which reduces the loss of small expressions.
- Multi-feature is extracted in the feature extraction process by considering both low-level and high-level features of the facial images to increase the accuracy of facial expression recognition.
- For increasing the quality of images and identify the pose of the facial images we perform bi-level preprocessing by normalizing the images considering illumination and pose of the facial image. For illumination normalization grayscale algorithm is used to reduce the occlusion, contrast etc. For pose normalization geometry-based polar transformation algorithm is used to estimate the pose for obtaining accurate results.
- To reduce the false positive rate, cluster-based segmentation is performed by using Improved Fuzzy C-means clustering algorithm and detects the facial landmarks precisely.

3. Introduction

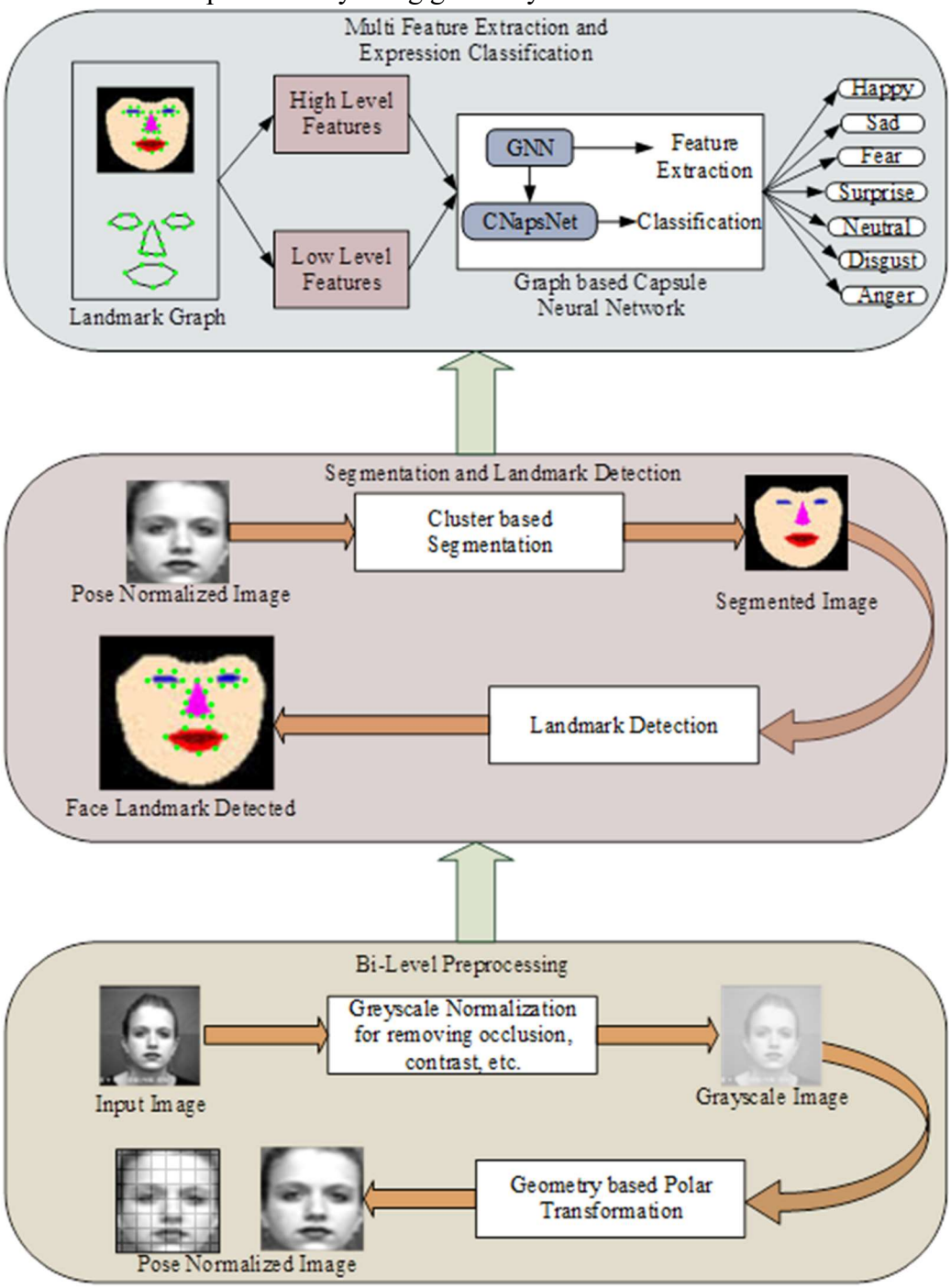
In this research, we have given importance to recognize facial expressions by using hybrid deep learning method. The accuracy of the facial expression is improved by performing preprocessing, segmentation, landmark detection, and feature extraction from the facial images. The important aim of this research is to develop expression recognition with high accuracy based on effective segmentation and classification. For facial expression recognition, we have taken CK+ (Cohn Kanade) and Japanese Female Facial Expressions (JAFFE) datasets. It has three sequential phases such as,

- Bi-level preprocessing
- Clustering-based Segmentation and Facial Landmark Detection
- Multi-Feature Extraction and Facial Expression Classification

3.1 Bi-level Preprocessing

We have taken the facial image as input from the given dataset i.e. CK+ dataset. Initially, bi-level preprocessing is performed for the facial images such as illumination normalization and pose normalization. In the illumination normalization occlusion, blurring effect; contrast, and shadowing

effect are reduced by using **Grayscale algorithm** to convert the normal facial image to grayscale image to eliminate these respective effects. After performing illumination normalization, facial angle and geometry are normalized by performing pose normalization by using **Geometry-based Polar transformation** to remove the background from the facial image and obtain frontal facial view for better facial expression recognition. The angle of the face is rotated within the image to get the frontal face pose by using polar transformation. At the same time, the position of the face is also important to recognize accurate facial expressions by using geometry transformation.



3.2 Clustering-based Segmentation and Facial Landmark Detection

After performing successful bi-level preprocessing, segmentation and facial landmark detection are performed. Segmentation of facial images is mainly performed to recognize the facial objects such as eyebrows, eyes, nose, mouth, and lips by using **Improved Fuzzy C-means clustering algorithm.** Traditional fuzzy c-means clustering algorithm is used for partitioning the images into segments but, it can't determine the center of the cluster. To overcome such drawbacks improved fuzzy c-means clustering algorithm is used to initialize the cluster's center by using **Firefly algorithm** along with fuzzy c-means clustering algorithm. After fulfilling the segmentation process, detection of facial landmarks is also performed effectively. 44 Feature points are used in the facial landmark for better facial expression recognition. 10 feature points are used for eyebrows (5 feature points for each eyebrow), 12 feature points for eyes (6 feature points for each eye), 9 feature points for nose, and 13 feature points for mouth (7 feature points for lips).

3.3 Multi-Feature Extraction and Facial Expression Classification

Various features are extracted from the landmark detected image and it is classified into two levels such as high-level features and low-level features. The high-level features are extracted based on the facial objects such as eyebrows, eyes, nose, lips, and mouth and the corresponding features such as *eyebrow slant, eye size, eye spacing, pupil size, nose length, nose width, nose wrinkle, mouth openness, mouth width, mouth curvature, tight lips, and lips droop*. The low-level features are *shape, texture, and color*. These features are extracted and classified by using deep learning algorithms such as **Graph-based Capsule Neural Network (GCapsNet).** The traditional capsule neural network has high complexity for large number of data sets. To overcome this drawback we integrate the capsule neural network with Graph neural network. Large number of datasets is managed by using Graph neural network and the classification is performed by capsule neural network to select the optimal features to reduce the overfitting and increase the accuracy. Seven important facial expressions are classified depend upon the extracted features; they are *neutral, happy, sad, fear, disgust, surprise, and anger.* Evaluation of the proposed work is performed by considering the following performance metrics,

- Accuracy
- Precision
- Recall
- Confusion matrix
- Facial landmark detection error

4. Literature Survey

[1] The author proposed an approach to recognize the facial expression from the static facial image by using hybrid deep learning networks. Initially, the expressional features were extracted from the facial image by using spatial attention convolutional neural network (SACNN). In this neural network, VGG-19 is used as a spatial attention module which is used to extract the pixel-based features from the facial images. After extracting the features, landmark detection is performed based on geometry of the facial

image by using attention mechanism based on long short-term memory networks (ALSTM), and this mechanism was also used to estimate different landmark regions' importance. Here FER-2013, CK+, and JAFFE datasets are used for experiment. The author used SACNN for extracting the pixel-based expressional features from the facial images. However, SACNN needs huge trained images to extract features effectively that increase the complexity for classification of facial expressions. Here, Batch normalization layer is included in the SACNN to reduce the internal covariance shift but, it needs huge batch size to normalize the outputs of the convolution layer to produce better results and Lack of considering the size of batch leads to high false-positive rate. To discover the facial landmarks from the extracted facial features ALSTM was used and it is also used to evaluate the landmark region's importance adaptively. However, lack of considering the overfitting problem when exploring the facial landmarks that decrease the accuracy of the facial expression classification.

[2] In this paper, authors proposed an approach for recognizing the facial expression by using deep convolution neural network (DCNN) with local gravitational force descriptor. Initially, the local features are extracted from the facial image by using the local gravitational force descriptor. After extracting the local features, DCNN is used to classify the facial expressions by dividing it into two major branches. The first branch of DCNN is used to extract the geometric features such as curves, edges, and lines present in the facial images. The second branch of DCNN is used to extract holistic features and classification. The classification score of facial expressions is computed by using a technique named score level fusion. Here FER-2013, CK+, and JAFFE datasets are used for experiment. Geometric features such as curves, lines, and edges of the facial images are considered by using DCNN. However, these geometric features of the facial images are not enough to classify the facial expressions accurately that leads to high false-positive rate. Here, DCNN is used for both feature extraction and classification of facial expressions from the corresponding features but, DCNN felt difficult to classify the facial expression when the facial image contains rotation or some angle of tilt that leads to less accuracy and more layers were used for the three types of layers present in the DCNN. Especially the number of max-pooling layers present is 5. However, more max-pooling layers present in the DCNN increase the probability of losing the efficient features that leads to high false positive rate. Preprocessing process is performed to align the angle and pose of the facial image. However, lack of considering the illumination parameters leads to face occlusion, contrast, etc., which decreases the classification of facial expressions accuracy.

[3] The author proposed an approach for facial expression recognition (FER) by using hybrid deep learning algorithms with pose considering face alignment. Initially, the pose-guided face alignment method is used to decrease the intra-class difference present in the facial images by considering three basic steps such as target pose discovery, template generation, and target matching. Angular symmetry is used for redundant features elimination and selects the efficient template by performing clustering with K-means clustering algorithm. Hybrid deep learning algorithms such as CNN and RNN are used for extracting the facial features and VGG-16 and ResNet are used for classification of facial expressions. Here Oulu-CASIA, CK+, AR, and JAFFE datasets are used for experiment. The results from the pose-guided face alignment method were clustered by using K-means clustering algorithm. However, this clustering algorithm cluster the dataset in k number, and the entire k number of clusters

has a single cluster that leads to high complexity to retrieve the extracted features for classification which decreases the performance. In the hybrid deep learning algorithms, RNN also plays an important role in feature extraction. However, the process involved in RNN is difficult which increases the training complexity in classification of facial expression. ResNet is used for classify the facial expression from the extracted features but, the time duration of this network is high that affects the classification with high time complexity. Here, the features are extracted from the clustering output to classify the facial expressions but the clustering of facial images is not enough to extract the features accurately that leads to high false-positive rate.

[4] The authors proposed an approach to discover the facial expressions by using ensemble rule with deep learning algorithm. The face expression recognition algorithms are classified into two types such as feature-based algorithm and convolutional neural network-based algorithm. Initially, the facial landmark is extracted from the input image. Perform frontalization by using frontalization algorithm to manage the pose by rotating and manage the brightness of the image. Shortcut CNN is used to extract the features from the facial image and for FER classification adaptive exponentially weighted average ensemble rule was used. FER-2013, JAFFE, and CK+ datasets were used for experiment. In this paper, the landmarks are detected with 68 feature points to perform facial frontalization but, FER needs few landmarks, for instance, the landmarks of jaw are not important to classify accurate facial expression. Considering all facial landmarks for frontalization increases the classification complexity and decreases the accuracy. Here, frontalization algorithm is used to consider the brightness rotation and pose of the facial images but, lack of considering the geometry features increases the false positive rate. CNN is used to extract the features from the frontalized facial image. However, CNN doesn't consider the coordinate frames of the facial images that lead to less accuracy in recognizing the facial expression.

[5] The author proposed an approach to detect the facial landmark by performing semantic segmentation. Initially, the architecture of semantic segmentation was designed to segment the facial images by encoding them and feature maps of the facial images were extracted from the encoder and the feature maps are decoded by corresponding feature maps. Secondly, VGG-16 is used to extract facial landmarks from the feature maps and extract the features of the facial image. For classification of images, the feature maps are given to the softmax layer and classify the feature maps based on their weights. For experiment analysis authors created the dataset by own and used VGG-16 for training the facial images for classification of feature maps based on weights to create facial landmarks. However, VGG-16 takes more time to train the images and needs more bandwidth that affects the performance.

[6] Authors proposed an approach to identify the facial expression from the facial images by using deep convolutional neural networks. Initially, edge computing is involved to ensure the privacy of the facial images from the cloud. The generative adversarial network (GAN) is modified by including circular consensus to create CycleGAN to train the facial images effectively. Information of class constraints is also included in the CycleGAN to improve style conversion process. The normal

classifier of GAN i.e. discriminator is also modified by including an auxiliary expression classifier to classify the facial expressions efficiently. For experiment analysis authors used JAFFE, CK+, and FER-2013 datasets to identify the recognition rate and GAN is modified by adding circular consensus and an auxiliary expression classifier to create CycleGAN for better classification. However, GAN can't have the ability to predict the state of the facial image that leads to less accuracy.

[7] Authors proposed an approach to recognize the facial expression by considering the pose and identity invariant of the facial image using dynamic multi-channel metric network (DML-Net). Initially, the DML-Net is used to learn the local and global features from various facial regions by using parallel convolutional networks. To discover the pose and identity-invariant of the facial expressions joint embedded feature learning is used. End-to-End training is performed for the facial images to recognize the facial expressions with low FER loss and overfitting. Here BU-3DFE, Multi-PIE, SFEW 2.0., and KDEF datasets are used to evaluate the facial expressions. DML-Net is used to extracting the features and classification of facial expressions from facial images. However, it is not suitable to classify the facial expressions in unconstrained environments i.e. various datasets.

[8] Authors proposed an approach to identify the facial expression by using CapsField technique. It is a technique with the combination of convolutional neural network (CNN) and capsule neural network. Initially, CNN is used to extract the features from the pre-processed array of facial images. Capsule network is used to route the facial features to select the features hierarchically. These methods reduce the redundancy of the features and make the classification process easier with effective classification of facial expressions. Here authors proposed the dataset named light field faces in the wild (LFFW) for experimental analysis. CNN was used with capsule network to extract the features and classify the facial expressions. However, CNN does not consider the alignment and position of the facial images that leads to less accuracy.

[9] Authors proposed an approach to recognize facial expression by using frequency neural network (FreNet). Initially, the facial images are pre-processed by considering rotation correction; ensuring the eyes are on horizontal line, and resizing using discrete Fourier transform (DCT). Block-FreNet is introduced to reduce the dimension and effective feature learning. The low-level features are extracted by using learnable multiplication kernel. After extracting the low-level features, high-level features are extracted by using a summarization layer. To classify the facial expressions from the extracted features ANN-based classification layer was used. FER-2013 dataset was used for experimental analysis. DCT is used to perform preprocessing for the facial images considering resizing, rotation correction, etc., but lack of considering the geometrical features leads to high false-positive rate.

[10] Authors proposed an approach to identify the facial expression by considering the geometry aware pose invariant using deep learning mechanisms such as generative adversarial network (GAN). Initially, the poses from the facial images were based on angles, and according to facial expressions, the facial landmarks were generated facial landmarks. The identity representation is detached by using

the facial shape geometry which was provided by the facial landmarks. GAN is used to extract the features and perform classification for detecting the facial expressions from the extracted features. Here BU-3DFE, SFEW, and Multi-PIE datasets were used for performing experimental analysis. GAN is used to extract the features from the facial images and perform classification for detecting the facial expressions. However, at the same time, GAN trains the generator and discriminator model that increases the training complexity.

[11] Authors proposed an approach to recognize the facial expression by considering the alignment and synthesis of facial images by using joint deep learning model. Initially, facial alignment is learned to generate the geometry code to perform facial synthesis. In the facial synthesis, three types of losses are considered such as discriminator loss, content similarity loss, and perceptual loss. Facial features are extracted i.e the expression and geometry code are extracted by considering the pose invariant and corresponding facial landmarks. After extracting the codes the facial expression is classified by using the softmax layer. The datasets used are BU-3DFE, SFEW, and Multi-PIE. Here, normalization is performed for the facial landmarks by considering the inter-ocular distance. However, lack of several illumination parameters such as shadowing effect, contrast, etc., leads to less accuracy.

[12] Authors proposed an approach to detect the racial landmarks by using heatmap offset regression technique. Initially, the regression network is divided into two stages such as structural hourglass network (SHN) and global constraint network (GCN). Preprocessing is performed to extract the features accurately. SHN is used to discover the facial landmarks at initial condition by using the heatmap. Improved inception ResNet is used in SHN for contextual feature representation learning. GCN is used to perform offset estimation for efficient landmark location. Loss function is proposed to improve the facial landmarks in a precise manner. Finally, the outputs from SHN and GCN are combined for accurate prediction of facial landmarks. The datasets used are 300W, AFLW, 300-VW, and COFW and heatmap offset regression is used to predict the facial landmarks in a precise manner. However, heatmap doesn’t adapt to the color of the images that leads to less accuracy.

5. Experiment and Discussion

5.1 Dataset

CK+ Dataset is used for facial expression recognition using deep learning. Cohn-Kanade (**CK+**) dataset having 593 images ranging from 18 to 50 years of age with both the genders and these images are divided into seven expression classes: anger, contempt, disgust, fear, happiness, sadness, and surprise. The facial image is taken as input from the CK+ dataset and the selected image loaded successfully as shown in Figure 1.

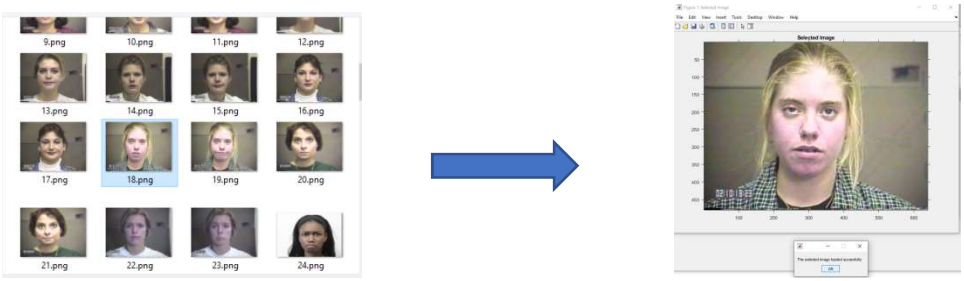


FIGURE 1. Loading of input facial image from dataset

5.2 Bi-level Preprocessing

In the bi-level preprocessing, we used the Grayscale algorithm for illumination normalization to convert the normal facial image to a grayscale image. After performing illumination normalization, facial angle and geometry are normalized by performing pose normalization by using geometry-based Polar transformation to remove the background from the facial image.

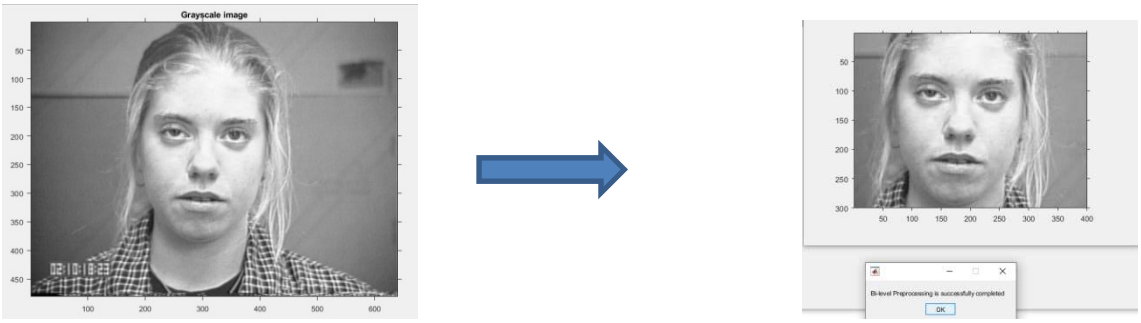


FIGURE 2. Bi-level Preprocessing

5.3 Segmentation

After performing Bi-level Preprocessing, the Clustering-based Segmentation and Facial Landmark Detection is performed. In this step, Segmentation of facial images is mainly performed to recognize facial objects such as eyebrows, eyes, nose, mouth, and lips. In this step we used Improved Fuzzy C-means clustering algorithm. Next, initialize the cluster's center by using the Firefly algorithm.

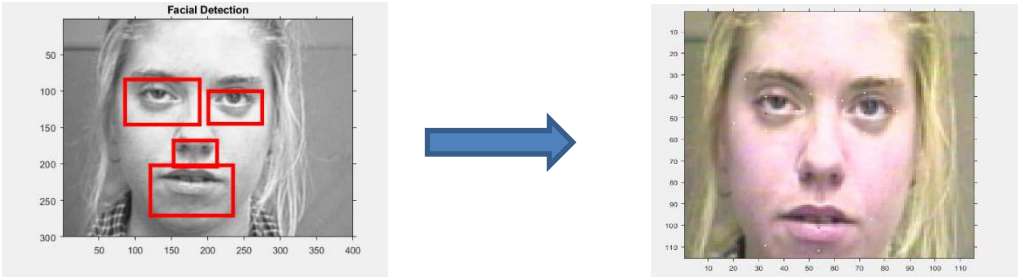


FIGURE 3. Facial Landmark Result

5.4 Feature Extraction

Multi-Feature Extraction and Facial Expression Classification: Various features are extracted from the landmark detected image and it is classified into two levels such as high-level features and low-level features. The high-level features are extracted based on the facial objects such as eyebrows, eyes, nose, lips, and mouth and the corresponding features such as eyebrow slant, eye size, eye

spacing, pupil size, nose length, nose width, nose wrinkle, mouth openness, mouth width, mouth curvature, tight lips, and lips droop. The low-level features are shape, texture, and color.

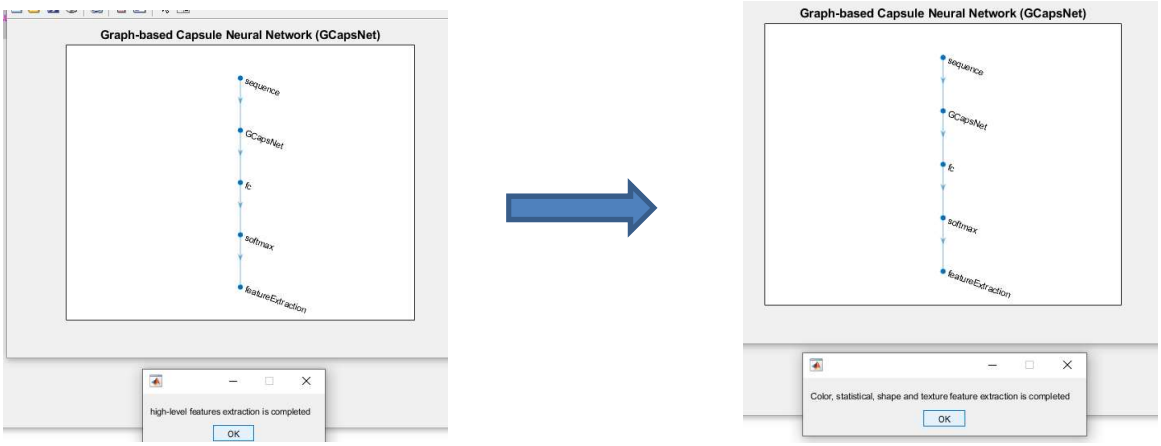
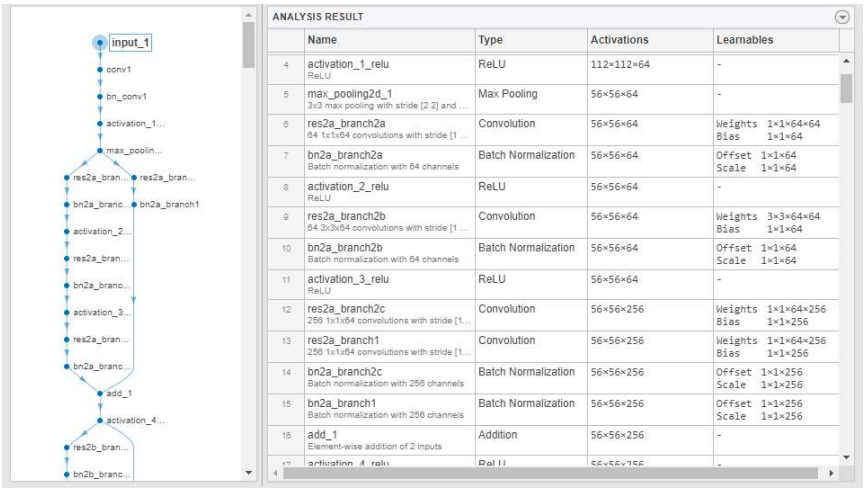


Figure 4. High level and Low level Feature Extraction

5.5 Classification

The deep learning network analyzer for classification is shown in the Figure 5 as follows:



	Name	Type	Activations	Learnables
4	activation_1_relu ReLU	ReLU	112×112×64	-
5	max_pooling2d_1 3x3 max pooling with stride [2 2] and ...	Max Pooling	56×56×64	-
6	res2a_branch2a 64 1x1x64 convolutions with stride [1 ...	Convolution	56×56×64	Weights 1×1×64×64 Bias 1×1×64
7	bn2a_branch2a Batch normalization with 64 channels	Batch Normalization	56×56×64	Offset 1×1×64 Scale 1×1×64
8	activation_2_relu ReLU	ReLU	56×56×64	-
9	res2a_branch2b 64 3x3x64 convolutions with stride [1 ...	Convolution	56×56×64	Weights 3×3×64×64 Bias 1×1×64
10	bn2a_branch2b Batch normalization with 64 channels	Batch Normalization	56×56×64	Offset 1×1×64 Scale 1×1×64
11	activation_3_relu ReLU	ReLU	56×56×64	-
12	res2a_branch2c 256 1x1x64 convolutions with stride [1...	Convolution	56×56×256	Weights 1×1×64×256 Bias 1×1×256
13	res2a_branch1 256 1x1x64 convolutions with stride [1...	Convolution	56×56×256	Weights 1×1×64×256 Bias 1×1×256
14	bn2a_branch2c Batch normalization with 256 channels	Batch Normalization	56×56×256	Offset 1×1×256 Scale 1×1×256
15	bn2a_branch1 Batch normalization with 256 channels	Batch Normalization	56×56×256	Offset 1×1×256 Scale 1×1×256
16	add_1 Element-wise addition of 2 inputs	Addition	56×56×256	-
17	activation_4_relu	ReLU	56×56×256	-

Figure 5. Deep Learning Network Analyzer

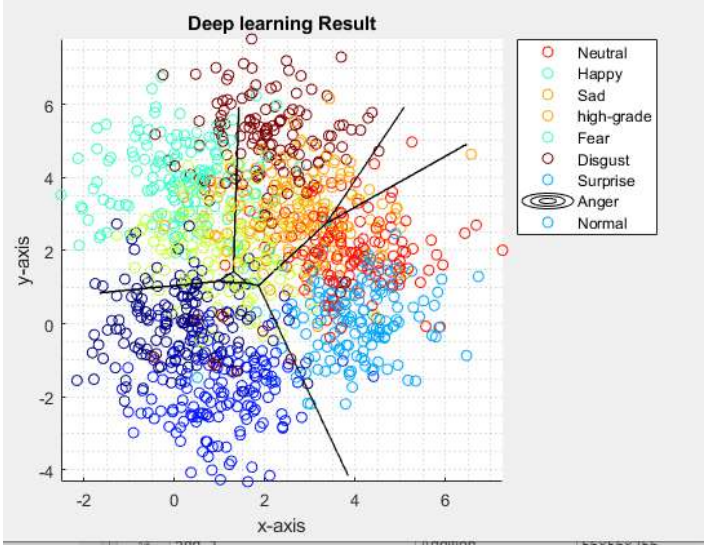


Figure 6. Deep Learning Result

The classification result, “ Sad” is shown in the Figure 7 as follows:

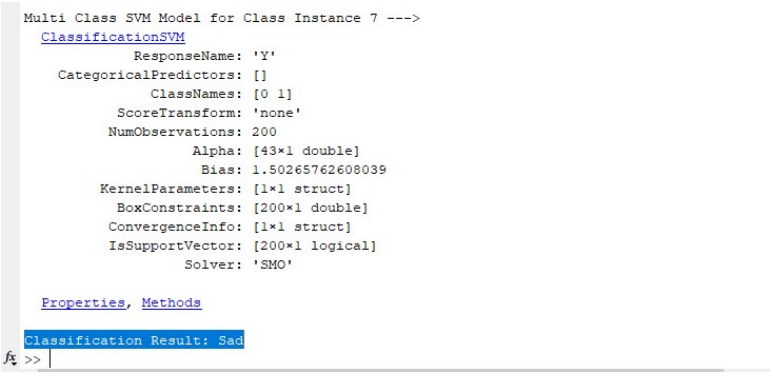


Figure 7. Classification Result

5.6 Performance Metrics

Finally, we evaluate the following performance metrics: Accuracy, Precision, Recall, Facial landmark detection error, Confusion matrix.

Figure 8. shows the Accuracy, Recall metrics while Figure 9. shows the Precision and Facial landmark detection error.

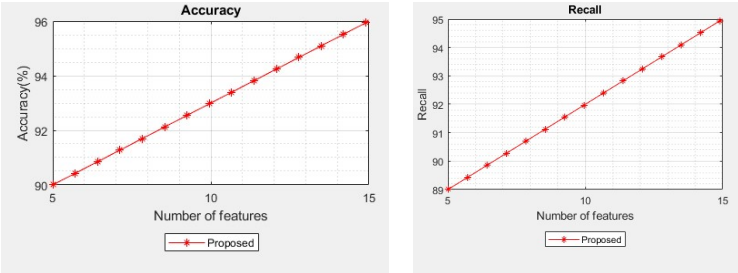


Figure 8. Accuracy, Recall

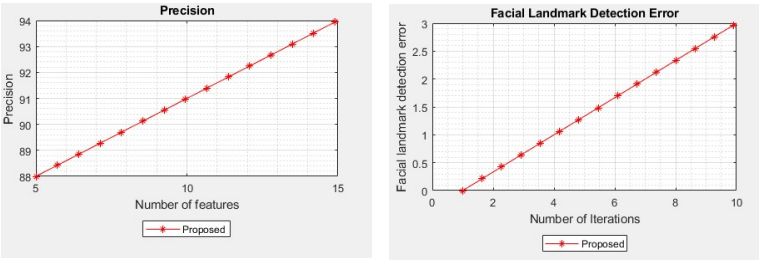


Figure 9. Precision, Facial Landmark Detection Error

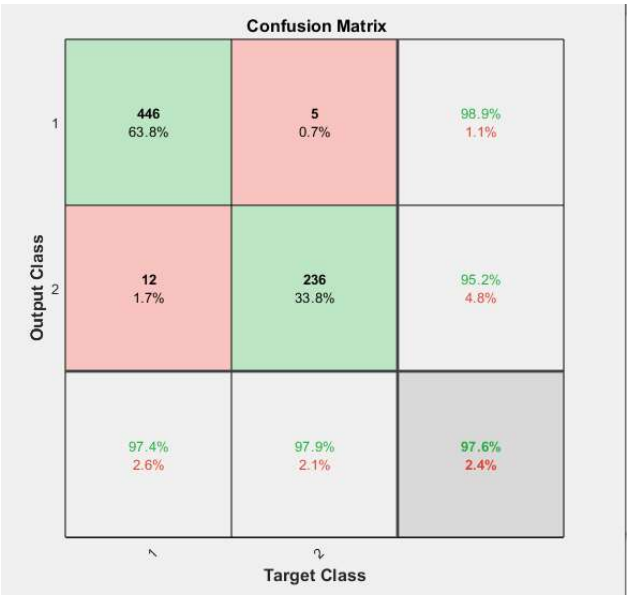


Figure 10. Confusion matrix

Conclusion

Facial Expression Recognition is used for classification of facial images into seven different expressions such as Sad, Fear, Anger, Happy etc. In this research, we evaluate the performance by using the following Performance Metrics such as Accuracy, Recall, Precision, Facial Landmark Detection Error, Confusion Matrix.

References

[1] Liu, C., Hirota, K., Ma, J., Jia, Z., & Dai, Y. (2021). Facial Expression Recognition Using Hybrid Features of Pixel and Geometry. *IEEE Access, 9*, 18876-18889.

[2] Mohan, K., Seal, A., Krejcar, O., & Yazidi, A. (2021). Facial Expression Recognition Using Local Gravitational Force Descriptor-Based Deep Convolution Neural Networks. *IEEE Transactions on Instrumentation and Measurement, 70*, 1-12.

[3] Liu, J., Feng, Y., & Wang, H. (2021). Facial Expression Recognition Using Pose-Guided Face Alignment and Discriminative Features Based on Deep Learning. *IEEE Access, 9*, 69267-69277.

[4] Tsai, K., Tsai, Y., Lee, Y., Ding, J., & Chang, R.Y. (2021). Frontalization and adaptive exponential ensemble rule for deep-learning-based facial expression recognition system. *Signal Process. Image Commun., 96*, 116321.

[5] Kim, H., Kim, H., Rew, J., & Hwang, E. (2020). FLSNet: Robust Facial Landmark Semantic Segmentation. *IEEE Access, 8*, 116163-116175.

[6] Chen, A., Xing, H., & Wang, F. (2020). A Facial Expression Recognition Method Using Deep Convolutional Neural Networks Based on Edge Computing. *IEEE Access, 8*, 49741-49751.

[7] Liu, Y., Dai, W., Fang, F., Chen, Y., Huang, R., Wang, R., & Wan, B. (2021). Dynamic multi-channel metric network for joint pose-aware and identity-invariant facial expression recognition. *Information Sciences, 578*, 195-213.

[8] Sepas-Moghaddam, A., Etemad, A., Pereira, F., & Correia, P.L. (2021). CapsField: Light Field-Based Face and Expression Recognition in the Wild Using Capsule Routing. *IEEE Transactions on Image Processing, 30*, 2627-2642.

[9] Tang, Y., Zhang, X., Hu, X., Wang, S., & Wang, H. (2021). Facial Expression Recognition Using Frequency Neural Network. *IEEE Transactions on Image Processing, 30*, 444-457.

[10] Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2020). Geometry Guided Pose-Invariant Facial Expression Recognition. *IEEE Transactions on Image Processing, 29*, 4445-4460.

[11] Zhang, F., Zhang, T., Mao, Q., & Xu, C. (2020). A Unified Deep Model for Joint Facial Expression Recognition, Face Synthesis, and Face Alignment. *IEEE Transactions on Image Processing, 29*, 6574-6589.

[12] Zhang, J., Hu, H., & Feng, S. (2020). Robust Facial Landmark Detection via Heatmap-Offset Regression. *IEEE Transactions on Image Processing, 29*, 5050-5064.

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Multimodal and Multilabel Genre Classification Of Movie Trailers

Aditya Kishore Jha
Information Technology
Delhi Technological University
New Delhi, India
adityakishorejha_2k18it006@dtu.ac.in

Akshat Batra
Information Technology
Delhi Technological University
New Delhi, India
akshatbatra_2k18it015@dtu.ac.in

Akshat Dubey
Information Technology
Delhi Technological University
New Delhi, India
akshatdubey_2k18it016@dtu.ac.in

Dinesh Kumar Vishwakarma
Department of Information Technology
New Delhi, India
dinesh@dtu.ac.in

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760773

***Abstract*—Movies are a diverse form of art and expressions. Unlike pictures and short clips, movies consist of a story-line which is deliberately made quite complex in order to engage the target audience. This paper presents evaluation of the usefulness of visual, textual and metadata-based functions for predicting the genre of a movie using movie trailers and analyzing it's visual features. The trailers were dissected into individual frames and were evaluated for key characteristics in order to divide them into different genres. Because previous articles employ an impractically large number of parameters to analyse the trailer, this approach aims to keep the number of parameters used to a minimum. The Moviescope dataset has been used which consists of about 5,000 movies with relevant information such as movie trailers, posters, plots and metadata. Linear Regression, KNN (K Nearest Neighbours) , Decision Tree, Random Forest and Artificial Neural Networks are just a few of the classification algorithms this research study has used and compared.**

***Index Terms*—movie trailers, movie genre classification, neural network, visual features, moviescope dataset**

I. Introduction

As equipments used for recording purpose and Internet connections are becoming more affordable and available, individualistic and professional video production and publicizing are becoming more accessible. Popular Online Streaming websites like Netflix, Hulu, Amazon now Occupy a large portion of the Internet bandwidth. Therefore, making video analysing techniques for these types of data becomes very important. Movie trailers have plot summary which is a brief understanding of a wide range of events which can be represented as a sequence of visuals.

There have been multiple works on compiling the videos for a variety of topics but most previous ones have focused on short clips rather than movies and isolated behaviours and patterns. At the same time, processing long hours of videos would require a very high computational power. Therefore, this paper proposes to use movie trailers instead of the movies because the two most important features they consist of are that they are like a summary of the whole movie capturing important aspects and they are widely available on platforms such as YouTube.

II. Related Work

There has been a numerous research in the field of genre predictions using different types of features extracted from trailers or Wikipedia. In this paper the focus is majorly on the prediction of genre using the text features and the video features of a trailer. One of the most recent and effective works in this field is performed in [1], where not only the features of videos and text have been analysed but the data set has also been created which consists of metadata from over 5000 movies with characteristics like movie trailer , with respective movie plot , movie poster. fastVideo and fastText have been used and it was stated that these despite their simplicity are well suited rather than LSTM-based representation for holistic classification task. The classification of genre in this paper is not unique, that is a single movie can be depicted in different class of genres, however this approach requires large amount of computing power as it extracts 200 frames from every movie trailer and each frame has 4096 features.

In the paper [2], multi-modal KDK classification for the movie trailers has been used. The model used to analyze the movie trailers was better in case of time and accuracy together with an accuracy of over 80 percent. The number of genres considered are four namely Action, Drama, Horror and Romance. High resolution frames have been used which can lead to more disk read and write time. The concept of YouTube-8M, a large-scale benchmark for video classification and representation learning is proposed in [3]. Different strong base lines for representation learning, containing deep learning and LSTM on the dataset and the results have been used which

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

show that these algorithm have edge over the popular used algorithms like logistic regression and SVMs.

The research paper [13] presented a framework in which features were extracted from scene analysis and the classification of genres was automated which focuses on extracting both the cognitive and affective features from the movie trailer. A dataset named EmoGDB was created, which contains 100 movie trailers tagged with popular movie genres and various sorts of induced emotions, as described in Paper [14]. The relationship between the emotions depicted in the movie trailers has been used together with their contributions towards genre classification problem. Similarly [4] has come across the fact that using low level visual features alone has a lesser impact for classification rather than using temporally-structured feature based on this intermediate level representation of scenes.The research paper [5] solely deals on the convolutional neural networks, and the design of the trained data-set, strength of the network architecture and good temporal modelling which are very crucial for analyzing generic videos. It has also been demonstrated that the used features can achieve results with simple linear model in low dimensional feature space and are cheap too.The [6,7] research paper demonstrate about the different algorithms used in the research carried out which are namely TF-IDF (Term Frequency Inverse Document Frequency) and CNN (Convolutional Neural Network). These algorithms are discussed in the forthcoming sections. The [9] research paper presented One-vs-Rest classifier for Multilabel classification.

In this paper the data-set known as moviescope has been analysed which consists of over 5000 movies data which has video features extracted from trailer videos on youtube and text features from Wikipedia and other features like about genre of the movie from IMDB. Although previous attempts have been made on video features on TV series and movies. Most of them just link video to another modality such as audio or poster features. Also data-sets present do not have large duration of videos and those who do have large videos do not have enough activities to encompass a whole movie that is why movie trailers are used in our research.

III. Methodology

A. Dataset

The dataset, Moviescope, consists of over 5,000 movie records collected from IMDB. This dataset was brought as part of a Kaggle Competition under Open Database license, and consists of a diversified collection of metadata information regarding movies included in this dataset and details regarding user interactions on social media. This dataset has been built up by crawling video trailers from YouTube for every movie in the dataset and text plots from Wikipedia.

Video trailers: For the collection of these clips, the most appropriate trailers on YouTube were looked upon, after which on launching a query the metadata from the top search result is extracted and the file is downloaded. To make the search even more accurate on the relevant queries the phrase “trailer” has been added to the end of the query, this makes the search spot

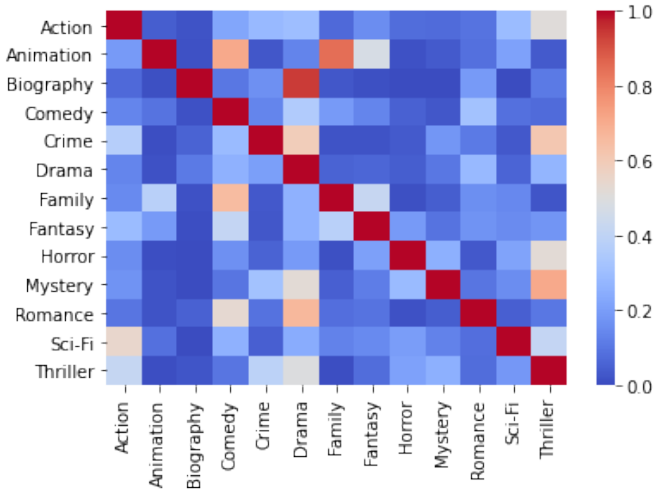


Fig. 1. Genre Correlation

on and the best result for the trailer of the related movie is searched. The data-set consists of around 4,996 movies using the method described earlier, and for the rest of the movies that were not traceable using this method the data was fetched manually. Most of the films that weren’t traceable were short films or from the earlier times.

Text plots: Each movie’s page on Wikipedia were looked upon and incorporated wherever the required page was accessible. The answer HTML was then parsed using reg-ex to extract the movie’s plot, which was then removed whitespace and Special symbol characters. Moreover the plot text was converted to lower case and using NLTK the stop words were removed from the text. However, if the searched movie on google didn’t have any relevant data on Wikipedia then the plot couldn’t be extracted, so the CMU Movie Summary Corpus dataset was used to locate the plot for these films. The above-mentioned dataset also contains plots from several movies, but without video features. Their IMDb page were looked upon and were checked for the suitable narrative for movies that aren’t even in this collection.

Metadata: The proposed collection is extremely diversified, containing films from 66 nations and 48 languages, as well as films from over 4000 different actors and over 2000 different directors from around the world. The most recent of the films in the dataset mentioned is from 2016 whereas the oldest one is from 1920. Action, biography, fantasy, animation, drama, crime, comedy, horror, sci-fi, family, mystery, romance and thriller are among the 13 movie genre genres. Larger number of categories and multi-label classification have been used here rather than single-label classification in comparison to earlier research. The correlation matrix for different labels in our dataset is depicted in Figure 1. For example, animated films are typically categorised as family films, although the contrary is less common. After looking at the raw data from this dataset it can be observed that the movies considered in each of the above mentioned genre, the genre with very few number of

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

movies is animation with just under 250 samples whereas the most occurring movie genre in this dataset is drama with over 2500 samples.

B. Data Preprocessing

It is a process in which the initial raw data is converted into an analyzable and understandable form on which various operations can be performed. For video features, 200 individual frames were extracted from one particular movie trailer. The individual frames were then used to extract feature vectors using the pretrained VGG-16 convolutional neural network model. Furthermore, the number of features were trimmed to almost a third of itself using dimensionality reduction techniques like PCA.

Here, 3454 movies have been considered from the dataset. Each movie has 10000 Text features . First Plot was cleaned and any punctuations were removed. All the text was then converted to lowercase. Stop words are removed from text using NLTK[10] . 80 % of the data is used for training and the rest is used for testing. The models used for classification are Logistic Regression and Random forest.

C. VGG-16 Convolutional Neural Network

VGG16 is a CNN model[8] which was first introduced in the paper "Very Deep Convolutional networks for Large-Scale Image Recognition" by K. Simonyan and A. Zisserman(Oxford University). It achieved an accuracy of 92.7% on being trained on the Imagenet Dataset which is a database of over 14 million images comprising of around 1000 classes. It replaces large kernel-sized filters with many kernel 3x3-sized filters in a row that is more advanced than AlexNet. AlexNet used large kernel-sized filters having 11 in the first convolutional layer and 5 in the second layer. VGG16 uses the NVIDIA Titan black GPU for which it was trained for weeks. The size of the input is fixed which is an RGB image of size 224 x 224. It uses multiple 3 x 3 convolutional layers and in one of the configurations it also uses 1 x 1 convolutional filters, to transform input channels linearly. The hidden layers have ReLu(rectification) non-linearity. It has a couple of disadvantages as well. It requires quite a lot of time to train and the weights of the network architecture are high. The weights of the network architecture are high and it requires a large amount of time to train. However, it is good for the purpose of learning and is quite easy to implement.

D. TF-IDF Vectorizer

TF-IDF as the name suggests stands for Term Frequency(TF) - Inverse Document Frequency(IDF). This is a widely used approach for converting text into a meaningful numerical representation, which is subsequently used by supervised machine learning algorithms to produce a prediction. Count vectorizer calculates the frequency of a word in relation to a vocabulary index, whereas tf-idf considers the weight of words across all publications.

$$tfidf(t,d,D) = tf(t,d) \times idf(t,D) \tag{1}$$

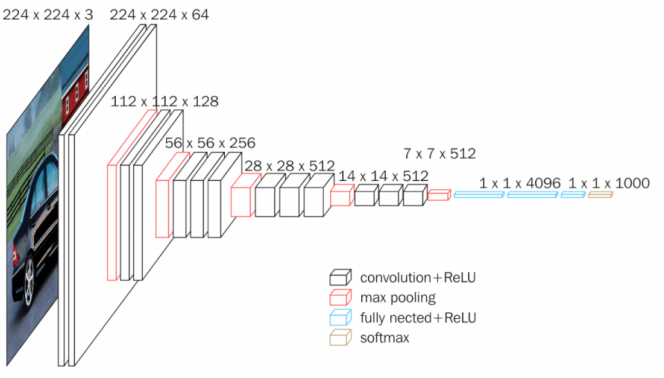


Fig. 2. VGG-16 Architecture

Where d stands for each document; D stands for the collection of documents ;t stands for terms.

We can define the term-frequency as counting function:

$$tf(t,d) = \sum_{x \in d} fr(x,t) \tag{2}$$

where fr(x,t) is defined as:

$$fr(x,t) = \begin{cases} 0, & \text{if x = t} \\ 1, & \text{otherwise} \end{cases} \tag{3}$$

The idf is defined :

$$idf(t,D) = \log \frac{|D|}{1 + |\{d \in D : t \in d\}|} \tag{4}$$

where $|\{d \in D : t \in d\}|$ stands for number of documents where term t is present,when $tf(t,d) \neq 0$ .Division by zero is avoided by adding one to the denominator.

E. Principal Component Analysis

PCA (Principal Component Analysis), is a method which is useful for dimensionality-reduction which is used to reduce the dimensions of a huge dataset. It transforms a large set of features into a smaller set at the same time retaining as much information as possible.

The accuracy is somewhat reduced but it is a trade off between accuracy and simplicity. Smaller datasets are easier to work on, analyze and faster for machine learning algorithms to learn from. Kernel PCA projects non-linear dataset into a higher dimensional feature space using a kernel function. At higher dimensions the dataset becomes linearly separable. It is very similar to the idea of Support Vector Machines. On performing dimensionality reduction, combined features vector size reduced from 14096 to around 1400.

F. Classification Algorithms

Multiple classification algorithms were used for predicting the genre using both text and video features including Logistic Regression, Naive Bayes, Decision Tree, KNN (K Nearest Neighbors), Random Forest and ANN(Artificial Neural Network). Logistic Regression is majorly designed for cases

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

TABLE I
VIDEO FEATURES RESULT

Model	Action	Anim	Bio	Comedy	Crime	Drama	Family	Fantasy	Horror	Mystery	Romance	Sci-fi	Thriller
Gaussian Naive Bayes	56.4	66.3	58.1	61.7	43.3	48.7	60.6	43.6	57.1	65.6	64.4	52.5	47.3
KNN	57.4	83.4	98.5	71.1	88.9	74.2	72.4	67.8	97.2	89.5	87.0	66.4	67.8
Decision Trees	54.7	76.4	96.1	66.2	79.8	65.3	65.1	56.8	92.9	82.5	78.8	56.9	61.4
Random Forests	60.4	83.4	98.5	70.9	88.9	76.2	72.8	68.9	97.2	89.5	87.0	66.8	69.3
ANN	64.4	84.2	97.2	71.3	84.0	68.8	74.9	67.6	96.0	86.4	81.4	68.1	62.5

where outcomes are in binary format. Decision Tree is used to handle regression and classification issues and is one of the most prominent supervised machine learning techniques. KNN is another supervised machine learning algorithm and is also used for both classification and regression problems and is generally used for pattern recognition. Random forests implements decision trees to solve the problems like the regression and classification problems by using various kinds of algorithms. The model in this case is trained for a set of data that is created as a heap using decision tree to maximize the accuracy. ANN can store information on the entire network and has the ability to work with incomplete information.

Random forest and ANN performed better than the other algorithms which can be due to the fact that these algorithms perform better with complex and non-linear data in comparison to other algorithms. There is not much risk of over-fitting and they can efficiently handle large data-sets at the same time being more accurate. Also, the ANN model performs slightly better than Random Forest since here the models are dealing with textual and image data in the form of video frames and Neural Networks are also better at dealing with a large amount of data.

G. Experimental Design

Video Features: In the original Dataset there were around 5000 movie trailers each having 200 frames and each frame has 4096 features extracted using the VGG-16 Convolutional Matrix. For now, we have considered 500 movies from the dataset, taking 10 video frames from each. Each frame has 4096 features and there are a total of 5000 frames. For training we have taken 4000 frames and the rest 1000 for testing. We have also used PCA for dimensionality reduction reducing the 4096 features to just around 1400. The models used for classification are Gaussian Naive Bayes, KNN (K Nearest Neighbours), Decision Tree, Random Forest and ANN (Artificial Neural Network). One-vs-Rest Classifier is used to Enable these binary classification algorithms to be used for multi-label classification.

Text Features: For now, we have considered 3454 movies from the dataset. Each movie has 10000 Text features . First Plot was cleaned and any punctuations were removed. All the text was then converted to lowercase. Stop words are removed from text using NLTK. 80 % of the data is used for training and the rest is used for testing. The models used for classification are Logistic Regression and Random forest. One-vs-Rest Classifier is used to Enable these binary classification algorithms to be used for multi-label classification.

IV. RESULTS

A. Classification Results from video features

The accuracy for all the individual genres using above mentioned classification models are shown in the table I. As it can be seen in the Figure 3 graph the accuracy of animated movies is fairly higher than the rest of the genres which can be explained by the fact that animated movies are completely defined by the video contents and it should be the most easier to catch when it comes to predicting genres through video features. Also, the reason why action is the least accurately predicted genre can be its slight association with many genres instead of strong associations with a few of them as observed from the genre correlation matrix. Other genres with higher accuracy in Table 1 could be due to the reason that there are less data points of those particular genres in the data-set.

B. Classification Results from Text features

It was observed that text features work well as a parameter for genre prediction than the video features especially for some of the genres like mystery, fantasy and biography which are difficult to predict using visuals. A decent f1-score of 0.4304 was achieved. The threshold value here was 0.5 which implies that the probabilities greater or equal than this were classified as 1 and less than these were classified as 0.

C. Combined Results from Video and Text features

On combining the results derived from the text and the video features the accuracies for the individual frames are mentioned in table II. From the table it can be observed that in genre such as family, which consists of words like child, father, mother shows significant improvement. Apart from family, fantasy genre also show notable refinement as text features such as magic, supernatural combined with video feature fabricate a much better result.

V. COMPARISON WITH OTHER MODELS

Both time efficiency and accuracy have been taken into consideration. Some of the previous classifiers take only accuracy as the metric but don't consider the feasibility of such models. The dimensions and the number of frames from one particular movie trailer were reduced as it was observed that the majority of the frames were very similar to each other. When taking 4910 frames with each frame having 1488 features the artificial

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

TABLE II
VIDEO + TEXT FEATURES RESULT

Model	Action	Anim	Bio	Comedy	Crime	Drama	Family	Fantasy	Horror	Mystery	Romance	Sci-fi	Thriller
Gaussian Naive Bayes	69.3	66.3	64.2	54.1	60.9	53.0	60.6	63.2	57.5	61.6	55.4	58.3	51.2
KNN	78.4	94.5	94.9	64.7	83.4	54.0	88.0	85.1	87.0	90.4	73.5	88.5	73.4
Decision Trees	67.1	92.0	89.0	54.5	70.6	52.5	80.3	79.4	78.8	82.2	67.0	79.6	61.3
Random Forests	78.8	94.5	94.9	59.2	83.4	51.2	88.0	85.1	86.8	90.4	75.4	88.6	72.0
ANN	76.3	87.4	93.9	65.6	79.8	57.5	87.9	83.5	81.6	87.3	68.4	87.3	66.4

TABLE III
CLASSIFICATION REPORT

Features	Precision	Recall
Video	0.65	0.63
Text	0.88	0.79

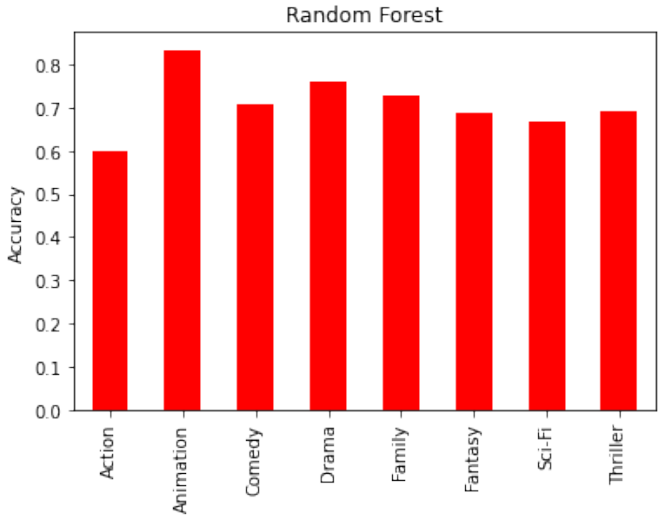


Fig. 3. Accuracy using video features

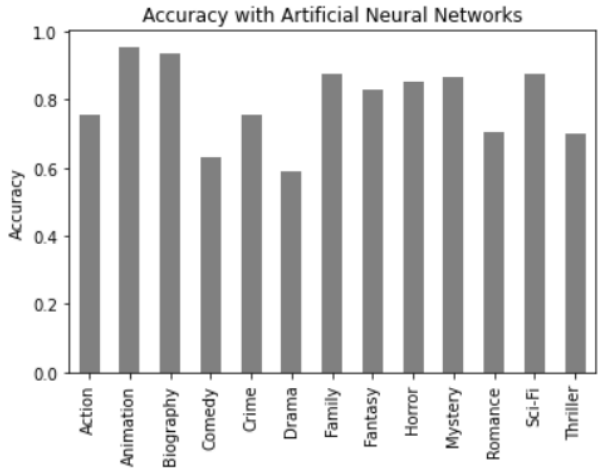


Fig. 4. Accuracy using combined features

neural network model takes about 3 minutes on average on a machine having Intel i7 10750H and DDR4 16GB ram. The model provides quite accurate results in different genre classifications considering the time efficiency and amount of features used. Combining the features of both trailer frames and text features also improved the performance of the model and then applying dimensionality reduction on the combined feature vector reducing the amount of features from around 15000 to around 1500 significantly reduced the time to train different models.

VI. CONCLUSION

The data-set used contains movie trailers, plots, posters and some metadata for about 5000 movies. It was found that despite the simplicity of video features, they are well suited for genre classification. Some particular genres were more accurately predicted by video features such as Animation and Drama and when combined with the textual features the performance of the model was significantly increased. In the future, classification based on posters and metadata would be looked at. Also, complex features such as the impact of videos on the audience, budget of a movie etc. using video features and metadata such as Facebook likes and reviews will be reviewed. Other Modalities such as Audio would also be introduced.

ACKNOWLEDGMENT

We would like to acknowledge everyone who played a part in our academic life and in making this project possible. We are very grateful to Dr. Dinesh Kumar Vishwakarma for lending us his supervision and guidance whenever we needed it. We also want to thank all of our department's faculty members for their leadership, persistent encouragement, and unwavering support for our project. In the end, we would thank our family and friends for their extended support throughout the project.

REFERENCES

[1] P. Cascante-Bonilla, K. Sitaraman, M. Luo, and V. Ordonez, "Moviescope: Large-scale Analysis of Movies using Multiple Modalities," Aug. 2019, [Online]. Available: http://arxiv.org/abs/1908.03180
[2] P. G. Shambharkar, M. N. Doja, D. Chandel, K. Bansal, and K. Taneja, "Multimodal kdk classifier for automatic classification of movie trailers," IJRTE, vol. 8, no. 3, 2019.
[3] S. Abu-El-Haija et al., "YouTube-8M: A Large-Scale Video Classification Benchmark," Sep. 2016, [Online]. Available: http://arxiv.org/abs/1609.08675
[4] A. del. Bimbo, S.-F. Chang, ACM Special Interest Group on Multimedia., and Association for Computing Machinery., MM '10: proceedings of the ACM Multimedia 2010 International Conference: October 25-29, 2010, Firenze, Italy. Association for Computing Machinery, 2010.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

[5] D. Tran, L. D. Bourdev, R. Fergus, L. Torresani, and M. Paluri, "C3D: generic features for video analysis," CoRR, abs/1412.0767, vol. 2, no. 7, p. 8, 2014.
[6] N. Srivastava, G. Hinton, A. Krizhevsky, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," 2014.
[7] C.-H. Huang, J. Yin, and F. Hou, "A text similarity measurement combining word semantic information with TF-IDF method," Jisuanji Xuebao(Chinese Journal of Computers), vol. 34, no. 5, pp. 856–864, 2011.
[8] M. ul Hassan, "Vgg16-convolutional network for classification and detection," Neurohive. Dostopno na: https://neurohive. io/en/popular-networks/vgg16/[10. 4. 2019], 2018.
[9] N. Fei and Y. Zhang, "Movie genre classification using TF-IDF and SVM," in Proceedings of the 2019 7th International Conference on Information Technology: IoT and Smart City, 2019, pp. 131–136.
[10] Jacob. Perkins, Python text processing with NLTK 2.0 Cookbook: over 80 practical recipes for using Python's NLTK suite of libraries to maximize your natural language processing capabilities. PACKT Publishing, 2010.
[11] A. Baig and V. Naik, "International Journal of Research in Advent Technology ACTION AND HORROR GENRE IDENTIFICATION OF MOVIES USING ARTIFICIAL NEURAL NETWORK," vol. 1, no. 5, 2013, [Online]. Available: http://www.ijrat.org
[12] K. Kowsari, A. Rupawala, D. Pujara, M. Shikalgar, E. Ukey, and B. E. Student, "IRJET-T weet Sent iment Analysis and St udy and Comparison of Various Approaches and Cla… IRJET Journal IRJET-DET ECT ION OF PHISHING E-MAILS USING DATA MINING IRJET Journal Text Classificat ion Algorit hms: A Survey Movie Genre Prediction from Plot Summaries by Comparing Various Classification Algorithms," International Research Journal of Engineering and Technology, 2020, [Online]. Available: www.irjet.net
[13] D. Kumar Vishwakarma, M. Jindal, A. Mittal, and A. Sharma, "Multi-level profiling of situation and dialogue-based deep networks for movie genre classification using movie trailers."
[14] A. Yadav and D. K. Vishwakarma, "A unified framework of deep networks for genre classification using movie trailer," Applied Soft Computing, vol. 96, p. 106624, 2020.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:00:06 UTC from IEEE Xplore. Restrictions apply.

Journal of Xi'an Shiyou University, Natural Science Edition ISSN : 1673-064X

My Land My Will - A Novel Blockchain Based Land Registry System

Rohan Vashist
Dept. of Electrical Engineering
Delhi Technological University
New Delhi, India
rohanvashist920@gmail.com

Saumya Swarup
Dept. of Electrical Engineering
Delhi Technological University
New Delhi, India
swarupsaumya65@gmail.com

Sneha Sood
Dept. of Electrical Engineering
Delhi Technological University
New Delhi, India
snehasood116@gmail.com

Anamika Chauhan
Asst. Professor, Dept. of Information and Technology
Delhi Technological University
New Delhi, India
anamika@dtu.ac.in

Anupama
Asst. Professor, Dept. of Electrical Engineering
Delhi Technological University
New Delhi, India
anupamaverma.91@gmail.com

***Abstract*— In India, land registration has always been a tedious task involving a horde of people, starting from the seller and buyer to the middlemen. The process of land registration is decades old, costly, time consuming and in the form of paperwork. Discrepancies in documents, such as counterfeit titles and a complete loss of the record, were extremely prevalent. Such circumstances result in costly judicial fights between opposing parties. Since the digital era has arrived, there have been attempts to digitalize and centralize this system, but issues like security and transparency are still exist. A blockchain is a distributed, immutable ledger that makes recording transactions and tracking assets easier. The immutability, auditability, and traceability of blockchain are luring governments all over the world to use the decentralized technology in property registration. In this proposed solution, a blockchain-based land registration system is designed and developed using high-end technologies to accelerate the process of registration. An additional 'Will' feature for registered properties has also been added. It has been concluded that the developed system can be successfully used based on its experimental use. It is a good alternative to the current procedure as it is modern and can yield better results.**

Keywords—Land Registry, Blockchain, Ethereum, Smart Contracts, ReactJs

I. INTRODUCTION

Land registry in India as well as in many parts of the world is a very slow and cumbersome process. A lot of cases where someone claims another's land, transaction frauds occur, documents are lost, and many more such issues have also been observed. Due to the thousands of land records that must be maintained, the current scenario in land registration indicates that it is faulty and less reliable [1]. The ledgers are kept in the form of paperwork or on less secure database systems. There is a high risk of this paperwork being misplaced, databases being hacked, or data being tampered with. Because land registrations necessitate monetary transactions, there is a risk of corruption, which increases the number of fraud cases.

A blockchain is a distributed, immutable ledger that makes recording transactions and tracking assets easier. On a blockchain network, anything of value may be recorded and sold, lowering risk and costs for all parties involved. Blockchain technology can provide us with a much more secure and reliable system for maintaining thousands of land records, as well as reducing the chances of corruption and data hacking. Smart contracts offer a viable alternative to distributed databases (which is necessary for a country like India which has a large population, thus increasing the number of land records). Blockchain can allow one to digitalize the entire land registration process.

The Indian government has launched numerous projects to build a Smart India/Digital India, and this project is a step toward the simple and hassle-free digitization of land registry process. It provides transparency by eliminating the possibility of corruption by ensuring verified transactions between concerned authorities, buyers, and sellers, as well as the absence of illegitimate transactions. Because there is currently no such government-regulated platform, the idea behind developing this project was to make the land registration process more resilient and to have a national impact [2].

The aim is to design a blockchain based land registry system that will help in securely maintaining all the records. This system will solve the problems faced by all the parties during the land registration and will make the process of land registration resilient and decrease the cases of fraud in the process. The system will also have a feature for users to add a will for their registered properties stating their heir. This will help in easy transfer of property in case of the owner's demise without any discrepancies [3].

II. LITERATURE REVIEW OF CURRENT SCENARIO OF LAND REGISTRY IN INDIA

A. Process of land registry

The process of document registration, signing deeds, getting documents notarized, verification of such data, changes in ownership, and other property transactions are all referred to as land registration. These deeds include documents like

survey documents, property maps, property tax receipts, record of rights, etc. The current owner of a land is responsible for property's past ownership records to be kept in check during such transactions. This process in our country is taken care of by the concerned authorities of each state as

every state has its own property registration departments, having specific requirements concerning forms and registration charges [4].

Since these data records can be very large and distributed among various states, it is difficult to keep them up to date which results in discrepancies among the data. Though measures have been taken in the past to resolve this issue, the success rate was negligible. There were cases of redundancy in the data in terms of actual property boundaries as they were never properly set up on the maps [5].

The current process of land registration is mentioned below:

- Firstly, the seller and the buyer have a verbal agreement on the sale or purchase of the property.
- After the agreement of the verbal contract, it is written down on paper, which is accomplished by the creation of a sale paperwork.
- After the deed is prepared, it is essential to make the payment of pay stamp duty. It varies from state to state (3-7 percent)
- The next requirement is to carry out and register the sale deed by visiting the Sub-registrar of Assurances office and submitting all the necessary documents.
- The final step is to have the property's mutation recorded in the land and income records. This can be done by going to the local municipal government.

This procedure appears to be outdated and sluggish.

B. Issues in Land Registry

- Space Constraints: It is very difficult and time-consuming to maintain land records that are nearly a century old as all this paperwork requires a lot of space to be stored in. Therefore, digitization of all documents and information is necessary, as well making them available to the general public will be a step towards eliminating the current difficulties.

- Lack of uniformity: Different states have different languages which makes it difficult for people who are not familiar with the language to understand the contents of the documents. Thus, in addition to digitizing, records should be maintained in a unified language understood by all, such as English.

- Fear of destruction of records: Physical records have the risk of being torn, damaged, lost, stolen, misplaced, etc as a result of a force of nature such as earthquake, flood or a fire. Thus, the need to upload the originally verified documents on a designated software arises.

- Frauds: As mentioned above data can be easily destroyed. In the absence of proper proof of ownership, the land registry becomes more prone to frauds.

- Lack of single window Title verification: In India, land ownership is presumed based on a variety of documents and records kept by numerous government agencies, making title investigation a laborious and time-consuming procedure. For example, registered documents are at the registration department, cadastral maps at survey departments, etc. It is found that this data may not be up to date or consistent with each other or maybe missing, which makes verification almost an impossible task to achieve [6].

Blockchain is a completely transparent, decentralised and secure system. Therefore, It appears as the most suitable solution to the current problems in the land registry system [7,8].

III. BLOCKCHAIN AS A SUITABLE SOLUTION

Blockchain gives a better approach to the land registry process because of its following advantages -

- Viable alternative to distributed databases: The blockchain system is nearly impossible to hack and cannot be manipulated. It can also hold a massive amount of data which is a plus point as land records can be humongous.

- Accelerated Process: The current process can be old-fashioned and slow. To register a land or transfer property one has to fill forms, fill up the deed, sign it, get it notarized and stamped, then submit it and wait for the approval of this paperwork by the government authorities. With the proposed digitized system in this paper, no one will have to go through all this hassle and can simply do that anywhere, without physically appearing at any office. Blockchain has the ability to confirm genuineness, therefore land can be registered or transferred legally by the landowners without involving third parties for verification [9].

- Reduction in Frauds: Nowadays, it is very easy for any fraudster to fabricate data and impersonate title ownership digitally, using different software that can edit data. On the other hand the proposed application in this paper enables users to submit property relevant documents on the Blockchain network, where concerned authorities may sign and confirm it, and this can be used to prove your ownership and avoid any fabrication by retaining immutable records [10].

- Transparency: Blockchain can be used to track the ownership history of any piece of land. Loan or mortgage procedures can be time-consuming due to administrative challenges. Smart Contracts, on the other hand, aid in the simplification of this procedure by automating validated transactions. As a seller or buyer, one can use the proposed application to generate a digital, decentralized ID and complete the transaction. The registrar will first confirm the transfer, then payment will be made, ownership will be transferred to the new buyer, and the entire transaction will be recorded in the system, making the process smoother and faster than it is now [11,12].

IV. PROPOSED SOLUTION OF BLOCKCHAIN BASED LAND REGISTRY SYSTEM

The generic architecture for the proposed system is shown in figure 1. Suitable technologies used to facilitate the proposed work flow have been discussed below:

A. Technologies Used

- Smart Contracts: In this system, smart contracts are used as the backend, which is nothing but a simple programme that runs upon the meeting of certain conditions and is stored on Blockchain [13,14,15].

- Solidity: To write our smart contract we used solidity language (version 0.8.0) .

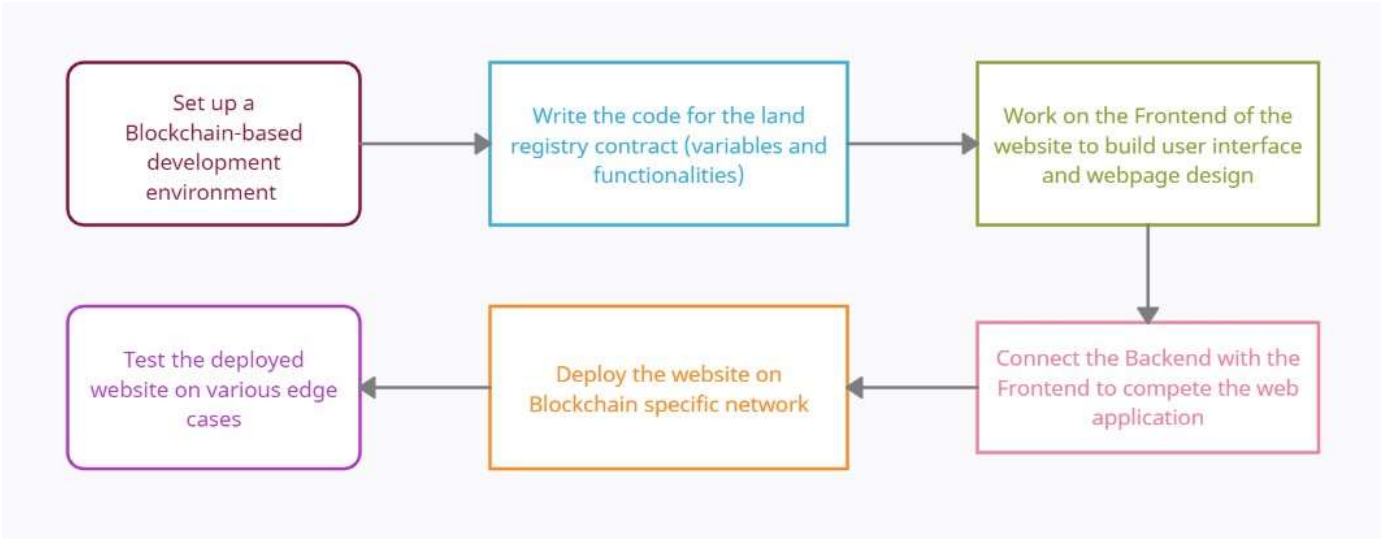


Fig. 1 Generic proposed architecture of the Blockchain based land registry system

- Ethereum: To run and deploy the proposed dApp, Ethereum blockchain is used. The smart contract was built and executed on the Ethereum platform, which provides an Ethereum virtual machine as the smart contract's runtime environment.
- Hardhat: Hardhat has been used as the development environment to compile, test, and debug the code on various test cases. It provides one with a blockchain server that runs on the developer's machines locally. It creates several false accounts with fake ETH tokens in order to perform transactions, and test the code without the need to perform these transactions on any public blockchain with real ETH tokens.
- Meta Mask wallet: It is a widely used cryptocurrency wallet that allows one to access their Ethereum wallet via a mobile app or an extension. It has been used to connect securely to this decentralized application, manage and store account keys, send and receive ETH-based tokens and currencies, and perform and broadcast transactions.
- ReactJS : To build a user side interface, ReactJs is used. It is a NodeJS framework that has been used as the development environment, because JavaScript provides extensive support for blockchain-based applications.

B. Design and Implementation

After carefully studying the process of land registry and taking all the current issues into account, an algorithm was devised to create a system facilitating land registry and Will creation using Ethereum Blockchain and ReactJS along with various other tools and libraries [16,17].

To define a smart contract, stakeholders have been identified by the developers and entities involved in the land registry and will creation process. All the roles are explained as follows –

- Government Official - A person who manages the whole system, verifies the sellers, buyers and all the information provided.
- Seller - A person who utilizes the system to sell properties and transfer land titles to buyers.
- Buyer - A person who wants to purchase land & utilizes the system to search for it, request access, & communicate with the seller in order to obtain possession of the land title.
- Land - The details of the property that a seller wants to sell are entered and are displayed to the buyer.

After defining the stakeholders and other entities, mappings were defined. The mappings are used to associate addresses with values. Mappings are like hash tables that can store key value pairs. For example, mapping a seller's address to its details etc. 17 such mappings have been defined:

- landsDetails
- GovtOfficialMapping
- SellerMapping
- BuyerMapping
- RequestsMapping
- RegisteredAddressMapping
- RegisteredSellerMapping
- RegisteredBuyerMapping
- SellerVerification
- SellerRejection
- BuyerVerification
- BuyerRejection
- LandVerification
- LandOwner
- RequestStatus
- RequestedLands
- PaymentReceived
- WillsDetails

Further, modifiers have been defined to control the behaviour of the functions such as restricting which stakeholder has the permission to call a particular function. For example, onlyGovtOfficial() modifier is defined, and if this is added to any function's definition, that function can only be called by the government official who acts as the admin. 8 such modifiers have been defined:

- onlyGovtOfficial
- onlyVerifiedLand
- onlyRegisteredSeller
- onlyRegisteredBuyer
- onlyRegisteredAddress
- onlyUnregisteredAddress
- onlyVerified
- paid

Finally, functions were defined for each step involved in the land registry and will creation process. The first function created is the constructor, which takes in the details of the government official, who acts as the admin of the website. The constructor is the first function called when the contract is deployed. 17 such functions have been defined that can be used to go through the complete process of land registry and Will creation:

- verifySeller
- rejectSeller
- verifyBuyer
- rejectBuyer
- verifyLand
- registerLand
- registerSeller
- registerBuyer
- updateSeller
- updateBuyer
- requestLand
- approveRequest
- landPayment
- ownershipTransfer
- createWill
- deleteWill
- transferWillProperty

To make the blockchain based land registry system easy to use for people of all age groups, a user interface was built for the system. ReactJS - a framework of NodeJs, was used for this. An admin page for the government official who acts as the admin of the website was created. Every request made by each user go through the admin first, who verifies every transaction going through the system. A separate home screen was also created for all the sellers and buyers which displays all the features available.

The EthersJS library is used to connect the land registry smart contract with the frontend react app. For that, a function connecting the forms and buttons in the frontend code to the functions in our contract was written.

The working of the website has been explained in the flowchart in figure 2.

C. Work Flow

First, a React App was created. Then a sample hardhat project inside the react app was created. After that, a smart contract named "LandRegistry" was written, defining the stakeholders' entities involved in the land registry process. Further created were mappings, variables and modifiers and functions for each step of the land registry and will creation step. Next step was writing the code for creating the user interface for thesystem using various libraries provided by ReactJS. After that, connection of the smart contract with the frontend application using the EthersJS library was established. And then the smart contract using the hardhat test network was deployed and tested the network against all the edge-cases.

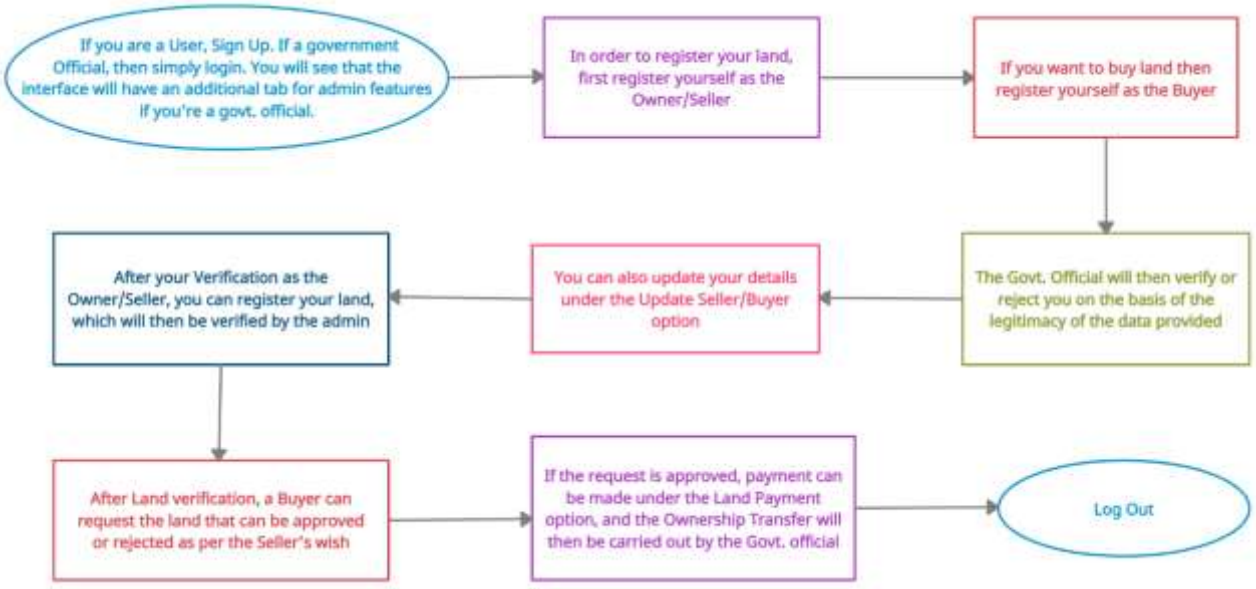


Fig. 2. Working of the website

V. Results and Discussion

The proposed system has been developed into a complete web application for a seamless land registry and Will creation process. Ethereum Blockchain was used as the secure database. A smart contract named "LandRegistry" was written along with mappings, modifiers and functions. The functions are defined to follow the land registry process step by step. The user interface was built using ReactJS and its different libraries. EtherJS is used to connect the backend part of the website to the frontend. Figure 3 shows the home screen of the website which is visible to all the users. A government official is made to act as the admin of the whole system. Every transaction goes through the admin first. Figure 4 shows the admin screen with all the functionalities present for the admin. Figure 5 shows the interface available to the user if they want to register as a seller. The other user functions also have a similar interface. Figure 6 shows how the transaction goes through metamask wallet. Lastly, figure 7 shows the console output "Seller Registered", from which its verified that the website works properly in all aspects. The complete working of the website has been explained in the flowchart in figure 2. [18].



3(a)



3(b)

Fig. 3. Home Screen



Fig. 4. Admin Screen



Fig. 5. Seller Registration

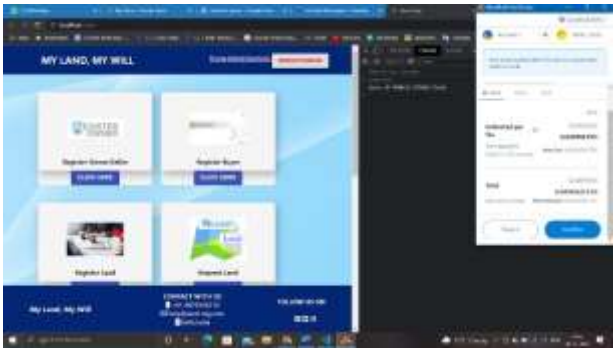


Fig. 6. Transactions using Meta mask



Fig. 7. Console Output showing seller is successfullyregistered

VI. Conclusion

In this paper, a better approach to the current land registry process has been proposed. The traditional land registry process has many issues, involvement of middlemen, discrepancy in documents, expensive and time consuming are to name a few. Blockchain has been used to build the proposed system. Using Blockchain a decentralized, transparent and secure system is built. A complete web application has been created using ReactJS for the frontend. By using the application, users can also add Will for their properties. It provides a smooth, hassle free and user-friendly experience, which makes the land registry process easy for people of all ages. The proposed system successfully eradicated almost all the issues in the current land registry system.

In future, wide range of payment options (for example- EMI) and modes (for example - Net Banking, UPI etc.) can be added to the system. Functionality for mortgage services can be integrated into the system.

REFERENCES

[1] Krishnapriya S, Greeshma Sarath, Securing Land Registration using Blockchain, Procedia Computer Science, Volume 171, 2020, Pages 1708-1715, ISSN 1877-0509

[2] https://www.ibm.com/in-en/topics/what-is-blockchain

[3] Vinay Thakur, M.N. Doja, Yogesh K. Dwivedi, Tanvir Ahmad, Ganesh Khadanga, Land records on Blockchain for implementation of Land Titling in India, International Journal of Information Management, Volume 52, 2020, 101940, ISSN 0268-4012

[4] https://www.business-standard.com/about/what-is-land-registry#collapse

[5] https://www.iifl.com/blogs/guide-property-registration-process

[6] https://blog.ipleaders.in/top-5-issues-maintaining-land-records-india-can-addressed/#:~:text=LACK%20OF%20UNIFORMITY%20AND%20POOR%20MAINTENANCE%20OF%20LAND%20RECORDS&text=Different%20states%20and%20different%20languages,the%20contents%20of%20the%20documents

[7] Ramya U.M., Sindhuja P., Atsaya R., Bavya Dharani B., Manikanta Varshith Golla S. (2019) Reducing Forgery in Land Registry System Using Blockchain Technology. In: Luhach A., Singh D., Hsiung PA., Hawari K., Lingras P., Singh P. (eds) Advanced Informatics for Computing Research. ICAICR 2018. Communications in Computer and Information Science, vol 955. Springer, Singapore.

[8] Fran Casino, Thomas K. Dasaklis, Constantinos Patsakis, A systematic literature review of blockchain-based applications: Current status, classification and open issues, Telematics and Informatics, Volume 36, 2019, Pages 55-81, ISSN 0736-5853, https://doi.org/10.1016/j.tele.2018.11.006.

[9] Svein Ølnes, Jolien Ubacht, Marijn Janssen, Blockchain in government: Benefits and implications of distributed ledger technology for information sharing, Government Information Quarterly, Volume 34, Issue 3, 2017, Pages 355-364, ISSN 0740-624X

[10] Laurie Hughes, Yogesh K. Dwivedi, Santosh K. Misra, Nripendra P. Rana, Vishnupriya Raghavan, Viswanadh Akella, Blockchain research, practice and policy: Applications, benefits, limitations, emerging research themes and research agenda, International Journal of Information Management, Volume 49, 2019, Pages 114-129, ISSN 0268-4012

[11] O. Ali, A. Jaradat, A. Kulakli and A. Abuhalimeh, "A Comparative Study: Blockchain Technology Utilization Benefits, Challenges and Functionalities," in *IEEE Access*, vol. 9, pp. 12730-12749, 2021, doi: 10.1109/ACCESS.2021.3050241.

[12] Chauhan, Anamika, Om Prakash Malviya, Madhav Verma, and Tejinder Singh Mor. "Blockchain and scalability." In 2018 IEEE International Conference on Software Quality, Reliability and Security Companion (QRS-C), pp. 122-128. IEEE, 2018.

[13] Chauhan, Anamika & Rishabh, & N S, Lokesh & Mittal, Pratham. (2022). A Deep Dive into Blockchain Consensus Protocols. 10.1007/978-981-16-4016-2_54.

[14] Zibin Zheng, Shaoan Xie, Hong-Ning Dai, Weili Chen, Xiangping Chen, Jian Weng, Muhammad Imran, An overview on smart contracts: Challenges, advances and platforms, Future Generation Computer Systems, Volume 105, 2020, Pages 475-491, ISSN 0167-739X

[15] Khan, S.N., Loukil, F., Ghedira-Guegan, C. *et al.* Blockchain smart contracts: Applications, challenges, and future trends. *Peer-to-Peer Netw. Appl.* **14,** 2901–2925 (2021). https://doi.org/10.1007/s12083-021-01127-0

[16] Soner, S., Litoriya, R. & Pandey, P. Exploring Blockchain and Smart Contract Technology for Reliable and Secure Land Registration and Record Management. *Wireless Pers Commun* 121, 2495–2509 (2021).

[17] Land Registration: Use-case of e-Governance using Blockchain Technology," *KSII Transactions on Internet and Information Systems*, vol. 14, no. 9. Korean Society for Internet Information (KSII), 30-Sep-2020.

[18] J. Golosova and A. Romanovs, "The Advantages and Disadvantages of the Blockchain Technology," *2018 IEEE 6th Workshop on Advances in Information, Electronic and Electrical Engineering (AIEEE)*, 2018, pp. 1-6, doi: 10.1109/AIEEE.2018.8592253

Computers and Electrical Engineering 101 (2022) 108071

Contents lists available at ScienceDirect

Computers and Electrical Engineering

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

Numerical investigation and temperature-based analysis of the analog performance of fully gate-covered junctionless FinFET☆

Gaurav Mangal, Aman Tyagi, Rishu Chaujar *

Department of Applied Physics, Delhi Technological University, Delhi, India

ARTICLE INFO

Keywords:
Junctionless transistor
Fully gate-covered FinFET
Short channel effects, temperature analysis

ABSTRACT

This work presents a temperature-based investigation of the analog performance parameters of a fully gate-covered junctionless SOI FinFET, which excludes source and drain extensions. The device is studied at 20/40 nm nodes with high-k gate oxide. The design provides a higher ON-current and better short channel characteristics (<70 mV/dec SS, around 30 mV/V DIBL) with an excellent switching ratio (100x better than conventional device). The temperature investigation reveals robust device integrity at higher temperatures. The short channel parameters degrade more rapidly for the conventional device with an increase in temperature as compared to fully gate-covered devices. At 600 K, the difference of DIBL was more than 25% between the conventional and proposed devices, the SS of the device increased by about 16.67%, and the switching ratio was better by an order of magnitude. This shows that the device is a potential candidate for applications where temperature resilience is of utmost importance.

1. Introduction

As the semiconductors industry continues on the path of Moore's law by doubling the components on-chip every two years [1], further increasing transistor density and the performance of a chip. The scaling of CMOS technologies and devices conventional devices has reached its fundamental limits to sub-nanometre regimes. This produces major challenges to technology which results from various undesirable issues with devices like leakage, power dissipation, and short channel effects [2,3]. These critical issues can seriously hamper and degrade the performance of these devices and circuits which must be addressed. Researchers have proposed numerous solutions and devices to alleviate these effects. Multi-gate architecture devices such as FinFETs are strong candidates as they scale better, have better power efficiency, reduced leakage currents, provide better mobility, and are overall robust at withstanding and more immune to these short channel effects as they provide better control over the channel region [4,5].

Besides that, even when these multi-gate FinFETs get smaller and more scaled down to be incorporated in a chip, defining these different PN junctions for source and drain is not an easy task, at such small scales. These nanodevices can be very sensitive to temperature change also. Temperature fluctuations, which can occur in such devices due to power dissipation and heating in the circuits, can also greatly increase the leakage current and affect other performance parameters as well [6]. As a result, Junctionless (JL) FinFETs are an attractive candidate due to their structural simplicity as they employ only a single type of doping throughout the channel, source, and drain regions eliminating the junctions, which has many advantages as employing these high doping gradients

☆ This paper is expanded from a conference paper presented at DevIC2021.

* Corresponding author.
E-mail address: chaujar.rishu@dtu.ac.in (R. Chaujar).

https://doi.org/10.1016/j.compeleceng.2022.108071
Received 23 June 2021; Received in revised form 28 April 2022; Accepted 4 May 2022
Available online 20 May 2022
0045-7906/© 2022 Elsevier Ltd. All rights reserved.

becomes extremely hard at these small lengths [7]. Due to this absence of PN junctions, junctionless transistors, when scaled down, offer more immunity against SCEs at such small scales. Moreover, JL FinFETs offer improved low-frequency noise behavior, better subthreshold, lower DIBL, and a reduced transverse electric field which in turn improves the device performance compared to an Inversion mode FinFET. These devices can achieve low leakage current in the OFF state and better on/off characteristics [8–10]. However, JL devices have high doping concentrations throughout the channel region which can lead to degradation of carrier mobility, which results in lower current drive and transconductance in comparison to IM mode FinFETs [11]. Rios et al. [12] also showed in their comparative study that IM transistors outperformed the JL transistors, but they suggested removing the source/drain extensions for more improvement of JL devices. Keeping that in mind, we studied our previously proposed fully gate-covered channel JL FinFET in which no S/D extensions are present, and we implemented two such structures, of different lengths, using this concept [13]. One, in which only the channel region is considered, and the rest source-drain portion is removed, Channel Only (CO)-JL FET, and the other, in which the remaining source-drain regions are also covered by extending the gate to cover the device entirely, Extended gate (EG)- JL FET [13].

In this paper, we performed extensive simulations to analyze the analog characteristics of our JL device which also incorporates a high-k dielectric, and compared its various parameters to a Conventional JL FinFET of similar dimensions. Temperature fluctuations can vary these analog parameters greatly, so it becomes imperative to analyze these at varying temperatures, which is necessary to better understand the device physics. Furthermore, such analysis of the temperature effect can highlight the additional applications beyond the room temperature, of a device. Thus, we further investigated the temperature effect of our devices for the various analog parameters of both the CO-JL FET and EG-JL FET. To the best knowledge of authors, no such reports have been published on temperature investigation of analog parameters of a Fully gate-covered channel employing a high-k dielectric for silicon-based JL FinFET. There are wide range of applications where such a device can be used especially in the field of military, satellite and deep-space communication, nuclear, aerospace industry, furnace temperature control as the temperature can vary to a wide range as the operating temperature varies from the nominal room temperature [14,15]. As for the applications, these devices act as basic building blocks of logic gates, RAMs, etc., with having much better characteristics than the traditional MOSFET and FinFET devices. It becomes thus crucial to gage the stability and reliability of devices for such wide temperature range applications. Further in analog circuits an infection or temperature-independent point where the effect of temperature compensates is very crucial to test the reliability as it is desired to bias the current at this point [16], which has been characterized in our study for a few parameters.

It is observed that the novel fully gate-covered structure improves the mobility of carriers as the devices have a larger area that is covered by the gate electrode, which causes carriers to face less resistance than they would have, in the source/drain extensions regions. Furthermore, the design owing to its larger gate electrode area and combined with high-k dielectric gate oxide HfO_2 makes it highly gate-controlled and reduces the SCE, to an extent that might occur due to removing the Source/Drain extension and has many advantages [17]. Even at high temperatures of 500 and 600 K, the device exhibits excellent performance when compared to the conventional device. Although packaging seems a bit challenging at these temperatures, the low subthreshold, leakage, and power dissipation of our device suggest better compatibility when compared to a conventional device [18]. Further, an additional heat spreader and heat sink can limit these challenges which can be explored in the future.

This paper is formated in the following manner. Section 1 contains the introduction and Section 2 which contains the details about simulation methodology and device structure. Further, Section 3 contains a comparative analog study between the devices and further temperature investigation of those analog parameters and finally, Section 5 presents the conclusion.

2. Simulation methodology and device structure

We have used the Silvaco TCAD EDA tools to simulate all the devices presented in this paper. Figs. 1, 2(a–c) show the modelled

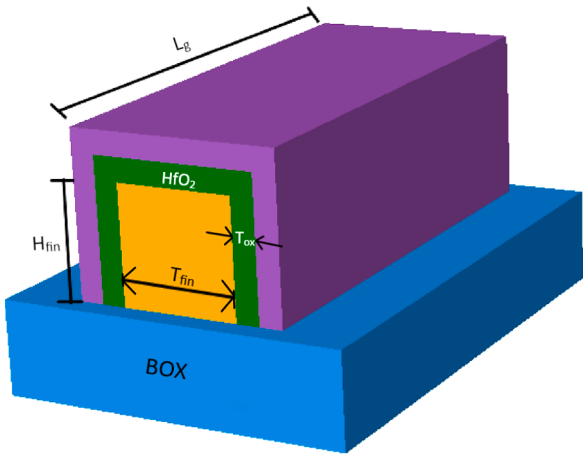


Fig. 1. CO-JL FET Structure.

devices and their cross-sections (along with their electron mobilities). The device structure parameters are shown in Table 1. The two gate-covered devices have the channel dimensions set as 20 and 40 nm, with Silicon as the channel material. These two devices are compared with a conventional JL FinFET with $L_g = 20$ nm. The channel doping concentrations of all the devices are set the same as $2 \times 10^{19} cm^{-3}$ n-type. The length of the S/D regions ($L_{S/D}$) in the conventional JL-FinFET is kept at 10 nm, whereas for the fully gate-covered JL FinFETs S/D regions are excluded. SiO_2 is used as the gate oxide for the conventional JL FinFET and HfO_2 is used in both fully gate-covered JL FinFET devices, with the thickness of the oxides, for all devices, at 1 nm. The fin height (H_{fin}) and fin thickness (T_{fin}) are both kept at 5 nm. A metal gate is used as the gate electrode, with a work function, $\varphi_m = 4.6$ eV [13].

Many physical models were incorporated in the simulation to accurately incorporate the device physics and performance under applied biases. The field-dependent mobility (FLDMOB) model and the Lombardi (CVT) mobility model are included to model the effects of the parallel and perpendicular electric fields on the mobility of the carriers [19]. The CVT model incorporates the transverse field-dependent, doping-dependent, and temperature-dependent parts of the carrier mobilities together, as shown in (1).

$$\mu_T^{-1} = \mu_{AC}^{-1} + \mu_b^{-1} + \mu_{sr}^{-1} \tag{1}$$

The above equation is known as Matthiessen's rule. The individual components signify the various components that affect the carrier mobilities. μ_{AC} describes the surface mobility which is bound by the scattering due to acoustic phonons. μ_b is the component of the carrier mobility that is affected by the scattering due to the intervalley phonons. Both these components rely on the lattice temperature, T_L, to accurately depict the scattering intensities of the carriers at the surface and within the substrate, respectively. The third component is the surface roughness factor which doesn't explicitly depend upon the lattice temperature [19]. Table 2 below shows the mobility models included that are explicitly dependent on T_L. Both The mobility models depict degradation of electron mobilities at higher temperatures.

The SRH recombination model models the phonon transitions in the presence of trap charges. The model also describes the relation of the recombination rate with the lattice temperature T_L, shown in Table 2. The first term in the denominator affects the recombination rate more, due to the exponent being positive. We could get an insight into how the recombination rate may get affected due to an increase in temperature. Increasing the temperature results in a denominator with a lower value and thus a higher recombination rate. Similarly, the Auger recombination model is also included, which models the Auger recombination of carriers, a three-particle transition of mobile carriers. The Fermi-Dirac statistics model is included instead of the default Boltzmann statistics model to obtain greater accuracy at small device dimensions. The inclusion of the Fermi-Dirac statistics is necessary in cases of high device doping, i.e., degenerate materials. The bandgap narrowing and nonlocal BTBT models are included for the incorporation of bandgap

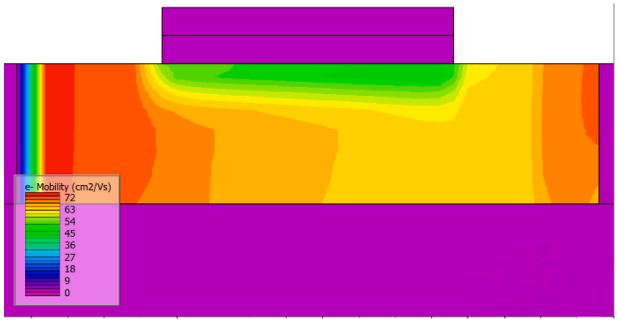


Figure 2(a)

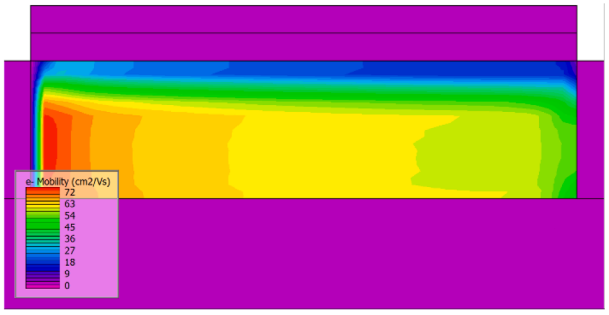


Figure 2(b)



Figure 2(c)

Fig. 2. Cross-section of the devices. (a) Electron mobility in the conventional JL Finfet. ($L_g = 20$ nm)(b) Electron mobility in the CO-JL FET ($L_g = 20$ nm). (c) Electron mobility in the EG-JL FET ($L_g = 40$ nm).

Table 1
Device structure parameters [13].

Parameters	Devices		
	Conv. JL-FET	EG-JL FET	CO-JL FET
L_g (nm)	20	40	20
T_{ox} (nm)	1	1	1
H_{fin} (nm)	5	5	5
T_{fin} (nm)	5	5	5
$L_{S/D}$ (nm)	10	–	–
N_D (cm^{-3})	2×10^{19}	2×10^{19}	2×10^{19}
ϕ_m (eV)	4.6	4.6	4.6

Table 2
Models for mobility and recombination [19].

Parameter	Equation (For electrons)
μ_{AC}	$\mu_{AC,n} = \frac{BN.CVT}{\left(\frac{E_\perp}{E_1}\right)EN.CVT} + \left(\frac{CN.CVT\left(\frac{N}{N_1}\right)^{TAUN.CVT}}{\left(\frac{E_\perp}{E_1}\right)^{DP.CVT}}\right)\frac{T_1}{T_L}$
μ_b	$\mu_{b,n} = MU0N.CVTexp\left(\frac{-PCN.CVT}{N}\right) + \frac{\left[MUMAXN.CVT\left(\frac{T_L}{300}\right)^{-GAMN.CVT} - MU0N.CVT\right]}{1+\left(\frac{N}{CRN.CVT}\right)^{ALPHN.CVT}} - \frac{MU1N.CVT}{1+\left(\frac{CSN.CVT}{N}\right)^{BETAN.CVT}}$
R_{SRH}	$\frac{pn - n_{ie}^2}{TAUP0\left[n + n_{ie}exp\left(\frac{ETRAP}{kT_L}\right)\right] + TAUN0\left[p + n_{ie}exp\left(\frac{-ETRAP}{kT_L}\right)\right]}$

narrowing due to high carrier concentration in the JL device and tunneling effects over the device characteristics [19,20].

The BQP model is used in the simulation of the devices, to incorporate the effects of quantum confinement to the bulk Silicon by associating a position-dependent quantum potential Q, which is added to the potential energy of the particular carrier, electrons in this case.

$$Q = \frac{-h^2}{2}\frac{\gamma * \nabla * M^{-1} * \nabla * (n^\alpha)}{n^\alpha} \tag{2}$$

The BQP model was calibrated with the Schrodinger-Poisson model, to give ngamma and nalpha as, 1.4 and 0.4, respectively. Other than these models, the Selberherr impact ionization model is also included along with the Hydrodynamic energy transport model. The Hydrodynamic transport model consists of two independent variables T_n and T_p that correspond to the carrier temperatures of the electron and holes, respectively. The model is more preferable as it treats mobilities and impact ionizations as functions of the

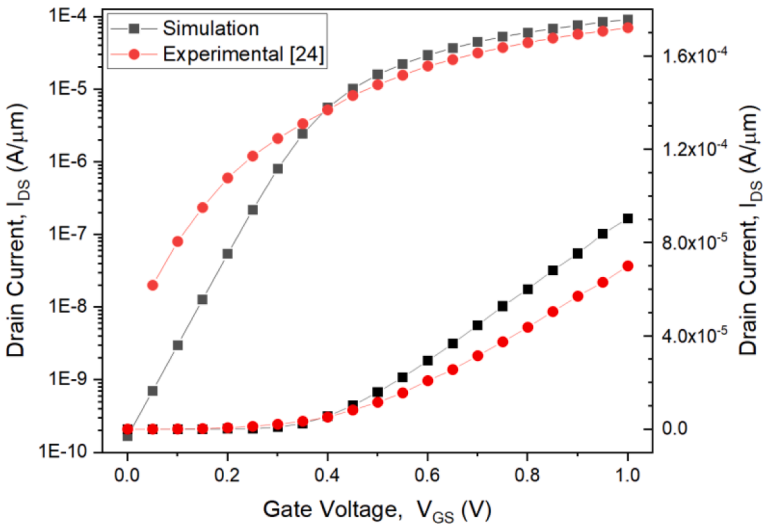


Fig. 3. Simulated results calibration with data from [12].

temperature, which is advantageous in such simulations.

The Block iteration method is used to numerically obtain the solutions as it is specifically used for a simulation involving the BQP model and energy balance models. To obtain the $I_D - V_{GS}$ device characteristics the gate voltage is ramped using steps of 0.05 V at a constant drain voltage of 0.5 V. Similarly, the I_D-V_{DS} curves are obtained at a constant $V_{GS} = 0.5$ V.

The models have been calibrated with the published results [12]. Fig. 3 shows a reasonable agreement between the simulation and the experimental data.

3. Results and discussion

3.1. Analog analysis

In this section, using the Silvaco Atlas TCAD simulator, we analysed the analog parameter and compared them for all three devices. We calculated current-voltage characteristics, transconductance (g_m), Output conductance (g_d), TGF, and Early voltage.

In Fig. 4(a), both the log scale and the linear scale graph of I_D - V_{GS} at a constant drain voltage of 0.5 V are shown. The three devices are compared and it is clear that the CO-JL FET which has a length of 20 nm has the highest ON-current value among the three [13]. The two main reasons, which can be highlighted, are due to fact that exclusion of the S/D regions reduces the resistance faced by the carries, and as the entire region is gate covered which causes more carriers to flow with less resistance and greater mobility in the channel region of the device, as shown in Fig. 2. The other reason, which can be highlighted, is the use of high-k gate oxide which also increases gate control of the device over the channel and helps in switching the device faster [17]. Similarly, the I_D-V_{DS} curves at a constant gate voltage of 0.5 V are shown in Fig. 4(b).

The transconductance (g_m) tells the switching speed of a device [17]. Greater the value of transconductance for a device, the greater its switching speed. It is also responsible for the noise performance of the device. The transconductance (g_m) can be given by (3), for such a device. Fig. 5(a) shows the plot of the transconductance against the gate voltage (V_{GS}). Again, it is evident from the graph that the CO-JL FET has the highest transconductance (g_m) value among the three devices [13], as transconductance depends on I_D and the CO-JL FET device has higher current values because of its better gate controllability and higher electron mobility.

The output conductance(g_d) of a device is given by (4). In Fig. 5(b), the plot of output conductance (g_d) of all the three devices is shown. For a device to have better analog performance, it is advantageous to have a value of the output conductance as low as possible, as a lower value is a good indicator of better DIBL and even modulation characteristics [21]. The EG-JL FET has the lowest output conductance amongst the devices shown in the graph which is due to the EG-JL structure having a longer channel length in comparison to both CO-JL and the conventional FinFET [13].

$$g_m = \frac{\partial I_D}{\partial V_{GS}} \tag{3}$$

$$g_d = \frac{\partial I_D}{\partial V_{DS}} \tag{4}$$

Transconductance generation factor (TGF) is another important analog parameter which is represented by (5) it shows how efficiently a current can achieve the particular value of transconductance. Higher efficiency devices have higher TGF values and better gain performance at lower voltages [21]. In Fig. 6(a), the plot of TGF against the gate overdrive voltage which can be given as ($V_{GT} = V_{GS}$ - V_{TH}) with keeping the drain to source voltage constant as 0.5 V. From the plot it is clear that the curve of CO-JL FET is highest

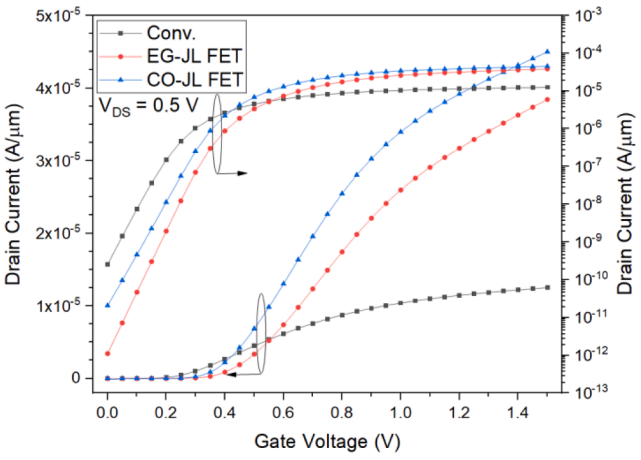


Figure 4(a)

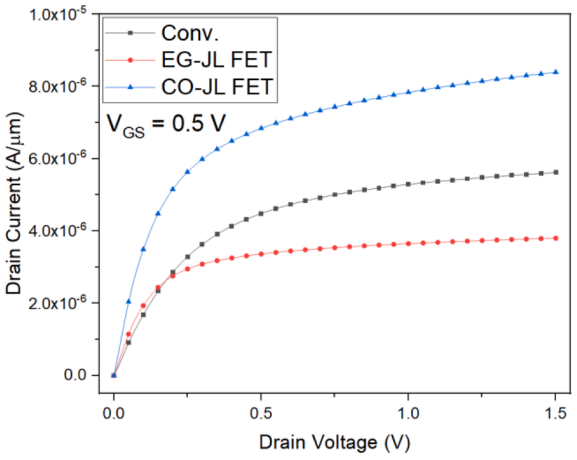


Figure 4(b)

Fig. 4. (a) Log and linear profile comparison of the I_D-V_{GS} curves. (b)Plot of Drain Current v/s Drain Voltage.

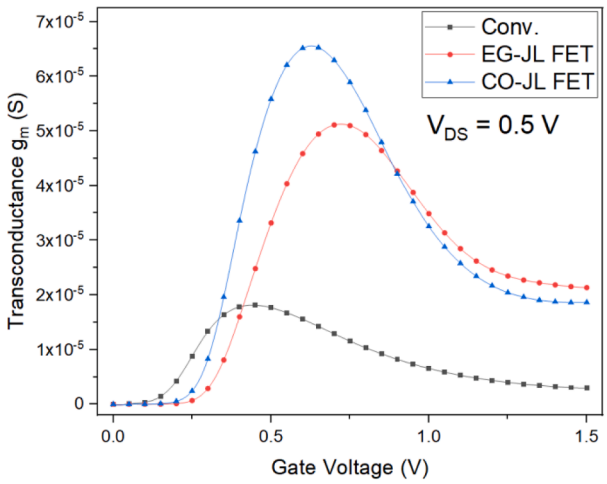


Figure 5(a)

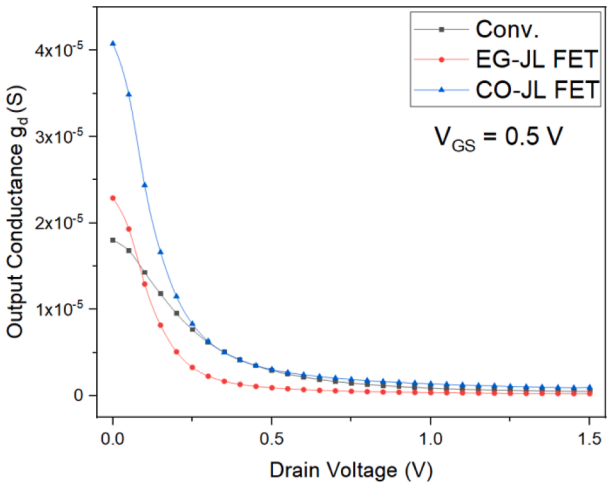


Figure 5(b)

Fig. 5. (a) Plot of Transconductance (g_m) v/s Gate Voltage (V_{GS}) for the devices. (b) Graph of Output Conductance (g_d) v/s Drain Voltage (V_{DS}).

among the three, and thus, has better gain performance. The CO-JL FET has the highest value of the transconductance (g_m) and thus has a higher value of the TGF parameter in comparison to other devices EG-JL and the Conventional JL FET [13]. The conventional structure has the lowest value amongst the three values for TGF, as it does not employ high-k dielectric.

$$TGF = \frac{g_m}{I_D} \tag{5}$$

Fig. 6(b) shows the collaborated plot of both Early Voltage (V_{EA}) versus V_{GT} and intrinsic gain (A_v) versus (V_{GT}) for all the mentioned devices. They can be defined by (6) and (7), respectively. Higher values of Early voltage are also much appreciated as they are a good indicator of short channel effects and better analog performance [21]. For both of these parameters, the value of EG-JL FET is maximum when compared to the other two devices [13]. Higher V_{EA} is due to the fact that the EG-JL FET has a g_d, which is significantly lower than others as seen previously, other than that the EG -JL FET also has fewer short channel effects as the length of the device is 20 nm longer than other two devices. Whereas the value for intrinsic gain is also highest for the same structure, due to the same reason.

$$V_{EA} = \frac{I_D}{g_d} \tag{6}$$

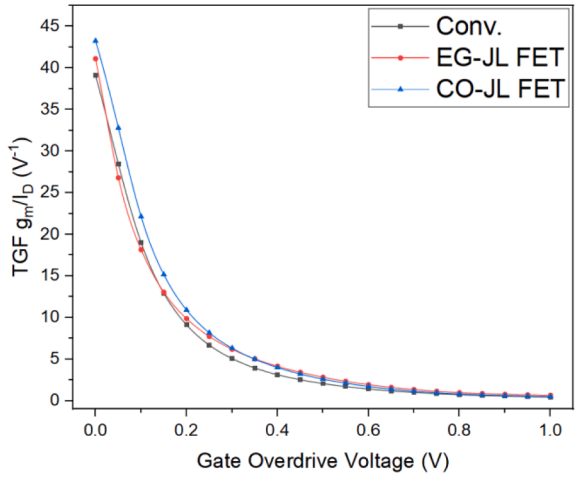


Figure 6(a)

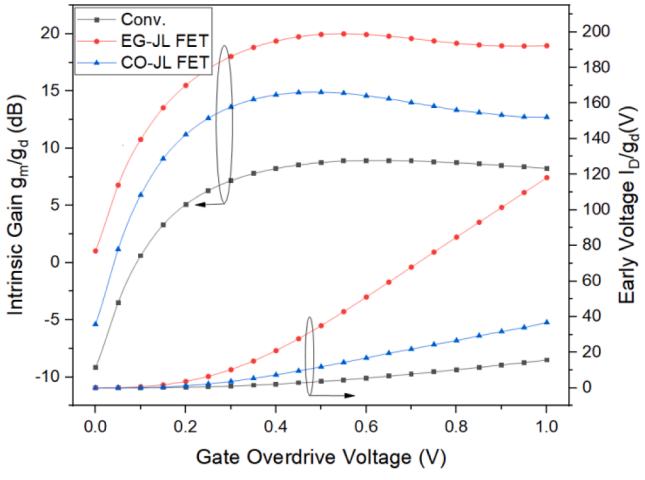


Figure 6(b)

Fig. 6. (a) Transconductance Generation Factor (TGF) v/s Gate Overdrive Voltage (V_{GT}). (b) Graph of Intrinsic Gain (A_v) and Early Voltage (V_{EA}) v/s Gate Overdrive Voltage (V_{GT}).

$$A_v = \frac{g_m}{g_d} \tag{7}$$

3.2. Temperature investigation

In this section, we analysed the temperature response of the analog parameters along with other important parameters like I_{ON}/I_{OFF}, DIBL, and subthreshold slope of both of our devices CO-JL FET and EG- JL FET.

Fig. 7(a) represents the threshold voltage (V_{th}) of the three devices and for any device, the threshold voltage is found to be inversely proportional to the temperature for which a general relation can be given as [22]:

$$\frac{dV_T}{dT} = \frac{d\phi_F}{dT}\left\{\frac{1}{C_{ox}}\sqrt{\frac{\varepsilon_0\varepsilon_{si}qN_{depl}}{\phi_F}} + 2\right\} - \frac{1}{C_{ox}}\frac{dQ_{ox}}{dT} \tag{8}$$

Where C_{ox}represents the capacitance of gate oxide, $\in_0$ is the vacuum permittivity, $\in_{si}$the relative permittivity of silicon, N_{depl} is ionized acceptor concentration, q is the charge Q_{ox} the oxide charge density, and φ_F is the Fermi potential.

Using (8), the reduction in threshold voltage can be explained as it varies significantly with temperature. Temperature increase reduces the bandgap and increases the intrinsic carrier concentration which can lead to the same amount of inversion at an even smaller gate voltage [16]. So, the threshold decreases with an increase in temperature. The temperature dependence of all the three devices can be seen ranging from 300 to 600 K in Fig. 7(b). It is thus shown that the threshold voltage of the EG FET is the highest for all the temperatures, among the three devices. V_{th} is significant for determining the ability of the device to switch ON and OFF.

Variation of Subthreshold Slope (SS) is also shown in Fig. 7(b). A generalized form can be given as in (9) for such MOS devices [23]: Fig. 8

$$SS = 2.3\frac{K_B T}{q}\left(\frac{1}{\lambda} + \frac{C_d}{C_{ox}}\right) \tag{9}$$

where the Boltzmann Constant k_B and T is the temperature, q the electronic charge, C_d the depletion capacitance and C_{ox} the oxide capacitance. The parameter $\lambda = 1 - (1/cosh(L_g/2l_c))$, where L_g is the gate length of the transistor and l_c is the characteristics length of the device. The graph can be explained by (9), the SS value increases with increasing temperature which results in degradation of performance of all the three devices. Moreover, SS is seen to be almost linearly proportional to the temperature. Fully gate-covered architecture has better subthreshold performance and the SS of conventional devices degrades rapidly specifically at higher temperature ranges over 400 K. The device thus shows better characteristics in the subthreshold regime as compared to the conventional. The fully gate-covered architecture, especially the EG-JL FET is better at withstanding the short channel effects.

To further discuss the device performance in the subthreshold regime the OFF-state leakage current or the subthreshold current is calculated again the relationship is almost linear with respect to the temperature on a logarithmic scale, the performance of all the devices degrades with increasing temperature as the subthreshold leakage current increases. However, even at high temperatures of 600 K, the performance of the fully gate-covered especially the EG-JL FET is better than the conventional device at 400 K. Performance deterioration for the fully gate-covered device with increasing temperature is much less when compared to the conventional device, which highlights the better subthreshold performance.

Fig. 9(a) represents the switching ratio of all three devices and I_{ON}/I_{OFF} is highest in the case of EG-JL FET due to the lowest I_{OFF} and

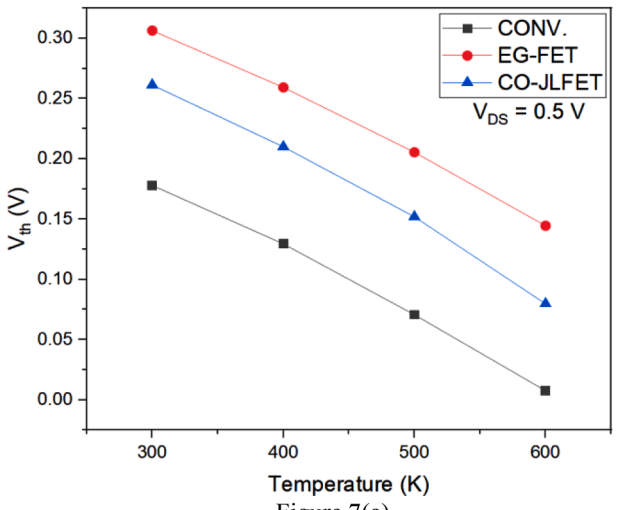


Figure 7(a)

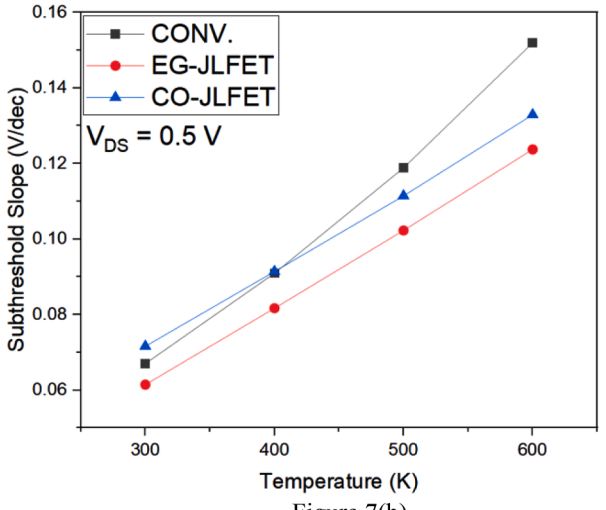


Figure 7(b)

Fig. 7. Variation of (a) Threshold voltage (V_{th}) with temperature for the devices. (b) Subthreshold Slope (SS) with temperature for the devices.

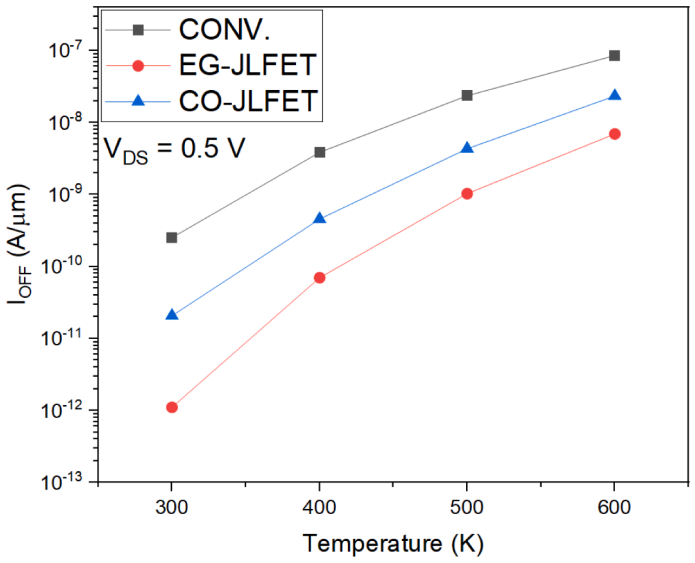


Fig. 8. OFF state current or subthreshold leakage current.

a considerably high I_{ON} [22], followed by the CO-JL FET. This shows that the novel architecture has a better response to leakage current along with a greater drive current and subthreshold characteristics, compared to conventional JL FET. It is also evident from Fig. 9(a) that as the temperature increases the switching ratio degrades which is due to the increase in leakage current (I_{OFF}) at high temperatures. The off-state current increases due to the fact that the diffusion current and the SRH recombination depend on the temperature, also they increase at high temperatures [24,25]. Also, at high temperatures, carrier saturation velocity reduces, which decreases the I_{ON} [16]. So, the switching ratio (I_{ON}/I_{OFF}) further degrades with temperature.

The Drain Induced Barrier Lowering (DIBL) phenomenon is strongly dependent on the gate length. In Fig. 9(b), the EG-JL FET, evidently, has the lowest value of DIBL, due to greater L_g and possibly due to the use of high-k gate oxide which provides better control of the channel. Similarly, in the case of CO-JL FET, the curve is seen to be better than the conventional device, with high-k oxide playing the major role here. But as the temperature increases, the potential barrier height reduces, at high temperatures, this can result in higher values of DIBL, degrading the device performance.

Fig. 10(a) and (b) represent the effect of variation of temperature on the drain current of the CO-JL and EG-JL FET, on the log and linear scale, respectively. It is clear that both the devices are better performing w.r.t. the conventional structure. At subthreshold voltages, the current is majorly due to diffusion and as a result, highly temperature-dependent [16], which is evident from both the graphs, since the current rises with increasing temperature at lower substrate biases. Other factors like the carrier concentration also increase, as the increment in temperature leads to the breaking of lattice bonds in silicon and the production of more electron-hole pairs [24]. But, as evident in both the plots 10(a) and (b), at higher gate voltages the drain current starts to degrade at higher temperatures, this is due to the fact that the as the temperature increases, the carrier mobility, and the drift velocity starts to degrade which in turn degrade the drain current performance of the devices and at a certain temperature the earlier increment due to diffusion is entirely compensated. This results in a sort of the point of inflection which can be seen in the graphs of both the devices.

Delay is an important factor to consider while implementing such semiconductor devices at the circuitry level. The delay depicts the overall speed of the device. Generally speaking, a faster device has a lower time delay. Device response times of a few picoseconds are very desirable for device implementations. The expression for device delay is given as in (10). In Fig. 11(a), (b), the delay performance for the CO-JL FET is reported to be better due to a higher drive current and a relatively low gate capacitance, although at higher temperatures it is degraded. whereas the EG-JL FET lags due to its longer gate length ($L_g = 40$ nm) offering greater propagation delay of the carriers, which is reflected in the expression of the delay in terms of the gate capacitance. The temperature response of both the devices is as expected as the delay increases due to increasing temperature.

$$\tau = \frac{C_{gg} V_{DD}}{I_{DS}} \tag{10}$$

Another important analog parameter that is analysed, is the effect of temperature on the transconductance(g_m) given by (3) for both the CO-JL and EG-JL FET which is shown in Fig. 12(a) and 12(b), respectively. Again, similar to the drain current the transconductance also increases due to carrier concentration increment at high temperatures and soon decreases at higher voltages as the carrier mobility is degraded. Transconductance is greatly dependent on the I_{ON} of the device and shows a similar point of inflection as in the I_D-V_{GS} temperature curves.

The output conductance (g_d) is analysed and shown in Fig. 13(a) and (b). In the previous section, it has been compared for all the three devices and here the temperature response has been shown for both the CO-JL and EG-JL FET in the plots. The output conductance also decreases as the temperature increases which gives a better current drive as the output resistance increase [24]. For

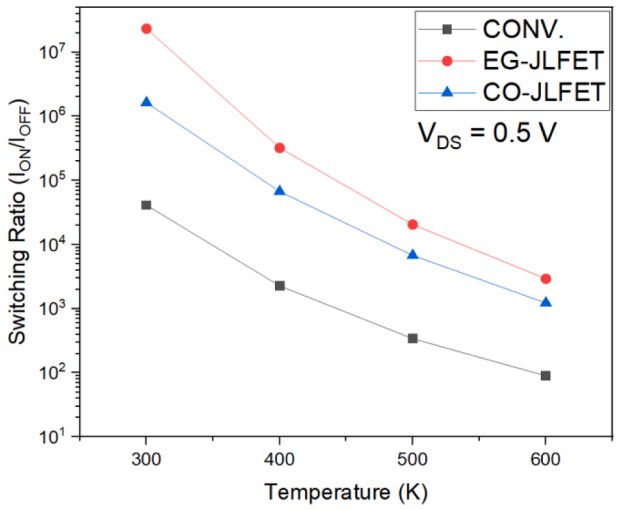


Figure 9(a)

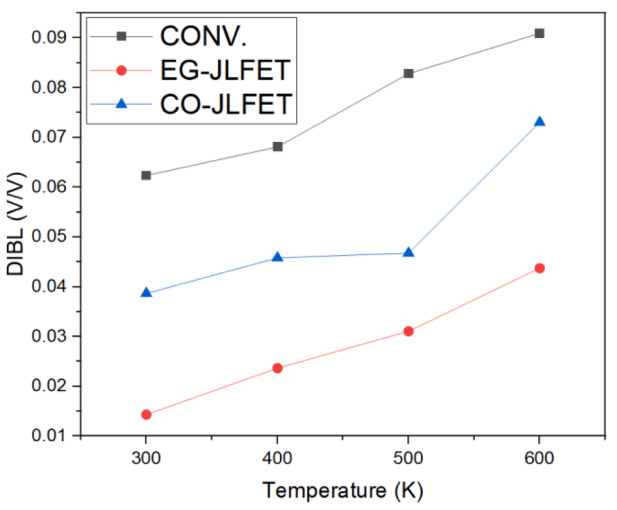


Figure 9(b)

Fig. 9. (a) Switching Ratio (I_{ON}/ I_{OFF}) variation with temperature. (b) Drain Induced Barrier Lowering (DIBL) variation with temperature.

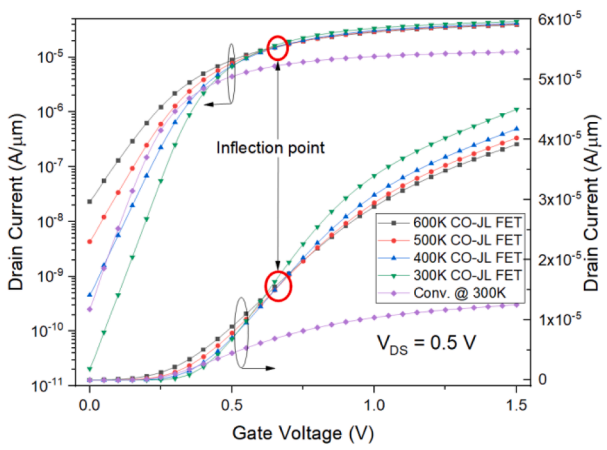


Figure 10(a)

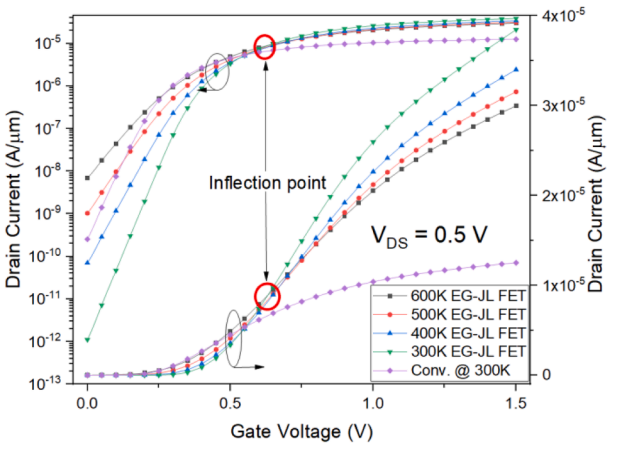


Figure 10(b)

Fig. 10. Temperature variation of drain current in (a) CO-JL FET (b) EG-JL FET.

both the graphs, it can also be seen that variation is negligible at higher voltages for the different temperature curves. The output conductance, however, has a better temperature response for the EG-JL FET, as at even 600 K it has a better response than the conventional device at 300 K as seen in Fig. 13(b).

Another parameter which is TGF has also been investigated for both the devices CO-JL and EG-JL in Fig. 14(a) and 14(b), respectively, for which a higher value is indicative of better analog characteristics. The TGF value decreases as the temperature increases. It is majorly determined by the I_D as it is more prominent in the subthreshold regions which result in lower values of temperature having higher values [16]. As the voltage increases the value with the lowest temperature still has the maximum TGF value but the temperature variation is close to negligible at these values which can be seen in the figures.

The intrinsic gain for both the CO-JL and EG-JL FET is shown in the collaborated Fig. 14(a) and (b) respectively and the intrinsic gain is very useful in evaluating the overall analog performance of the device. The value is higher for lower temperatures. But a similar trend like the g_m can be seen again as there is a point of inflection in both the graphs. This is due to the value of g_m itself being lowest at the subthreshold regions for low temperatures while g_d being always larger for lower temperatures. This creates a similar graph with an inflection point. Furthermore, as the value of g_m rises for lower temperatures, the value of gain for the lower temperature also becomes maximum. A summary of different performance parameters is shown in Table 3.

4. Conclusion

In this work, the analog performance parameters of the fully gate-covered Junctionless FinFET were analysed. The temperature

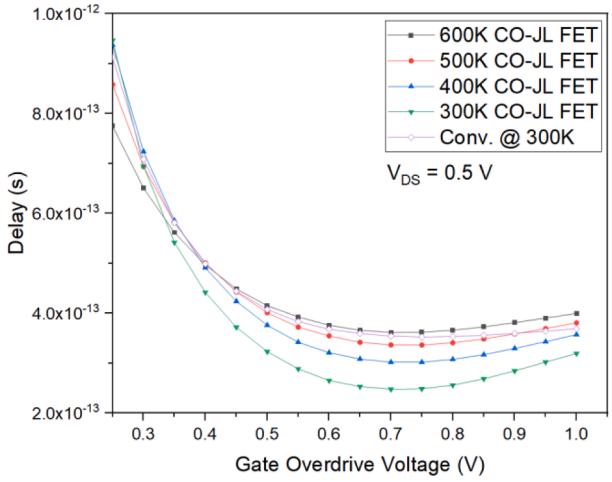


Figure 11(a)

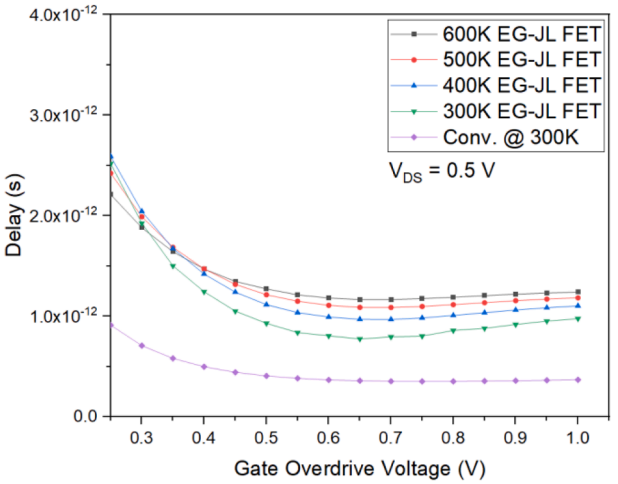


Figure 11(b)

Fig. 11. Graph of delay in (a) CO-JL FET with temperature (b) EG-JL FET with temperature, v/s gate overdrive voltage (V_{GT}).

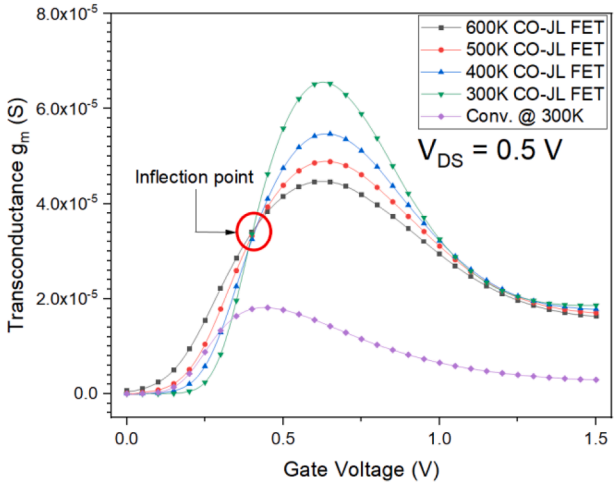


Figure 12(a)

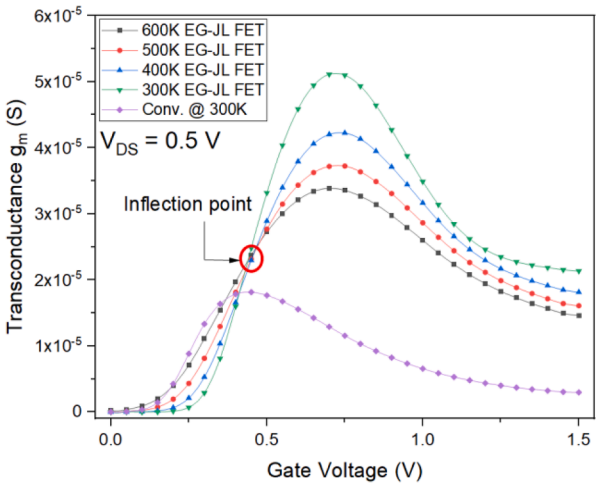


Figure 12(b)

Fig. 12. Variation of transconductance (g_m) in (a) CO-JL FET with temperature (b) EG-JL FET with temperature.

investigation of these parameters points to the overall better performance of the device architecture even at temperatures well above 300 K. The advantages of the fully gate covered structure are evident through the improvements of the device characteristics, g_m and V_{EA} increased by 3.6 times and 9.6 times, respectively for the EG-JL FET, as well as drain current increasing as much as 2.5 times for the fully gated structures compared to the conventional device. The degradation of the device characteristics at higher temperatures, due to mobility degradation and lattice bond breaking, impacted the device performance, but both the fully gate-covered devices presented a respectable advantage even at higher temperatures. The switching ratio of both the CO- and EG- FETs was greater than 10^3, as compared to 89 for the conventional device, at 600 K, which is ~10 times greater. The DIBL was reported to be ~20 mV/V below the conventional device. Furthermore, the SS degraded more rapidly for the conventional device than the fully gated structures, with increasing temperature. The analog performance parameters of the fully gate-covered devices at temperatures >300 K continued to stay better than the conventional device at 300 K. Max value of g_m, came to be 4.46×10^{-5} S and 3.39×10^{-5} S at 600 K, which for the conventional device was 1.81×10^{-5} S. The EG-JL FET also showed a better output conductance than the conventional device for all measured temperatures. Similarly, the TGF of the devices came to relatively same as the conventional device at 300 K, whereas gain was 50 and 200% more for the CO- and EG FETs. These parameters make the proposed device suitable for high current applications even at higher temperatures.

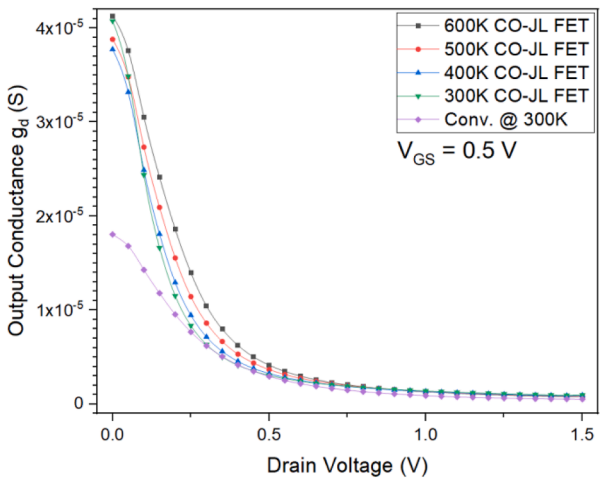


Figure 13(a)

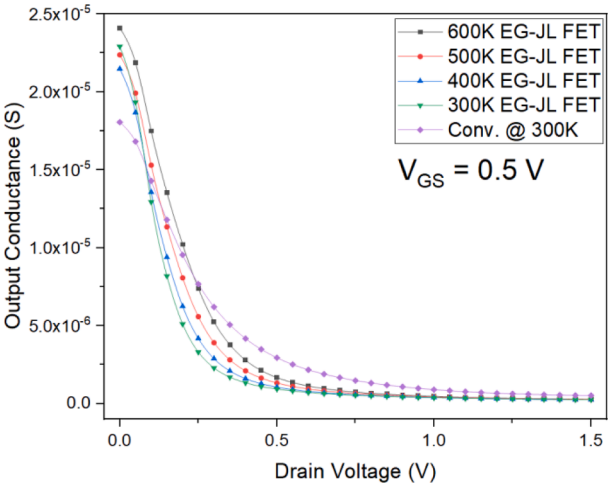


Figure 13(b)

Fig. 13. Output conductance (g_d) variation in (a) CO-JL FET with temperature (b) EG-JL FET with temperature.

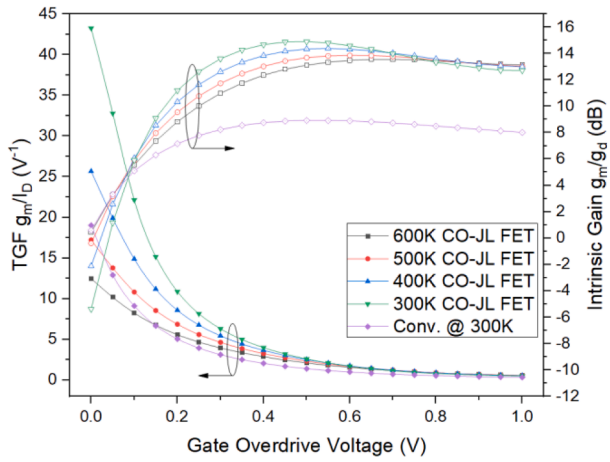


Figure 14(a)

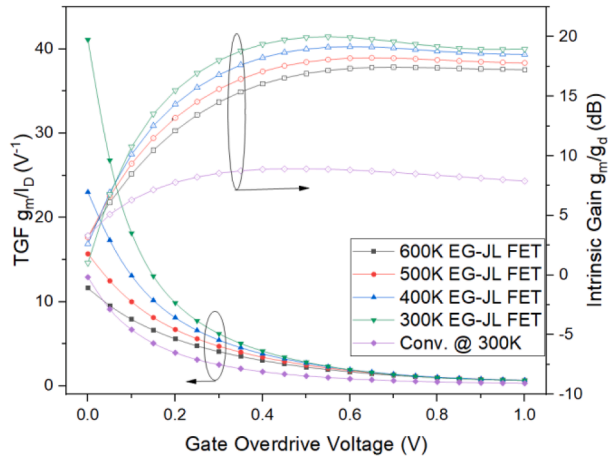


Figure 14(b)

Fig. 14. Variation of TGF and Intrinsic Gain for (a) CO-JL FET with temperature (b) EG-JL FET with temperature.

Table 3
Summary of different performance parameters.

Parameter	CO-FET @ 300K	CO-FET @ 600K	EG-FET @ 300K	EG-FET @ 600K	Conventional @ 300K	Conventional @ 600K
$I_{OFF}(A/\mu m)$	2.07×10^{-11}	2.34×10^{-8}	1.11×10^{-12}	6.97×10^{-9}	2.53×10^{-10}	8.53×10^{-8}
I_{ON}/I_{OFF}	1.64×10^{6}	1233.77	2.34×10^{7}	2941.92	41107.6	89.7819
$DIBL(mV/V)$	38.68	73.04	14.33	43.74	62.30	90.96
$SS(mV/dec)$	71.59	132.87	61.44	123.70	67.02	151.90
$V_{th}(V)$	0.2611	0.0797	0.3064	0.1444	0.1779	0.0076
$g_m(max)(S)$	6.53×10^{-5}	4.46×10^{-5}	5.11×10^{-5}	3.39×10^{-5}	1.81×10^{-5}	–
$TGF(max)(V^{-1})$	43.25	12.48	41.11	11.61	12.91	–
$Gain(max)(dB)$	14.90	13.55	20.01	17.45	8.92	–

CRediT authorship contribution statement

Gaurav Mangal: Investigation, Data curation, Conceptualization, Writing – original draft. **Aman Tyagi:** Methodology, Software, Visualization. **Rishu Chaujar:** Supervision.

Declaration of Competing Interest

The authors confirm that there are no known conflicts of interest associated with this publication and there has been no significant financial support for this work that could have influenced its outcome.

Acknowledgment

The authors would like to thank the Microelectronics Research Laboratory, Delhi Technological University, for supporting this work.

References

[1] Moore GE. Cramming more components onto integrated circuits. Proc IEEE 1998;86(1):82–5. https://doi.org/10.1109/JPROC.1998.658762. Jan.
[2] Frank DJ. Power-constrained CMOS scaling limits. IBM J Res Dev 2002;46(2.3):235–44. https://doi.org/10.1147/rd.462.0235. Mar.
[3] Chaudhry A, Kumar MJ. Controlling short-channel effects in deep-submicron SOI MOSFETs for improved reliability: a review. IEEE Trans Device Mater Relib 2004;4(1):99–109. https://doi.org/10.1109/TDMR.2004.824359. Mar.
[4] Hisamoto D, Lee WC, Kedzierski J, Takeuchi H, Asano K, Kuo C, Anderson E, King TJ, Bokor J, Hu C. FinFET-a self-aligned double-gate MOSFET scalable to 20nm. IEEE Trans Electron Devices 2000;47(12):2320–5. https://doi.org/10.1109/16.887014. Dec.
[5] Wann CH, Noda K, Tanaka T, Yoshida M, Hu C. A comparative study of advanced MOSFET concepts. IEEE Trans Electron Devices 1996;43(10):1742–53. https://doi.org/10.1109/16.536820. Oct.
[6] Semenov O, Vassighi A, Sachdev M. Impact of technology scaling on thermal behavior of leakage current in sub-quarter micron MOSFETs: perspective of low temperature current testing. Microelectron J 2002;33(11):985–94. https://doi.org/10.1016/S0026-2692(02)00071-X. Nov.
[7] Colinge JP, Lee CW, Afzalian A, Akhavan ND, Yan R, Ferain I, Razavi P, O'Neill B, Blake A, White M, Kelleher AM, McCarthy B, Murphy R. Nanowire transistors without junctions. Nat Nanotechnol 2010;5(3):225–9. https://doi.org/10.1038/nnano.2010.15. Mar.
[8] Park CH, Ko MD, Kim KH, Baek RH, Sohn CW, Baek CK, Park S, Deen MJ, Jeong YH, Lee JS. Electrical characteristics of 20-nm junctionless Si nanowire transistors. Solid State Electron 2012;73:7–10. https://doi.org/10.1016/j.sse.2011.11.032. Jul.
[9] Jeon DY, Mouis M, Barraud S, Ghibaudo G. Channel width dependent subthreshold operation of tri-gate junctionless transistors. Solid State Electron 2020;171:107860. https://doi.org/10.1016/j.sse.2020.107860. Sep.
[10] Kumar A, Tripathi MM, Chaujar R. Comprehensive analysis of sub-20nm black phosphorus based junctionless-recessed channel MOSFET for analog/RF applications. Superlattices Microstruct Apr. 2018;116:171–80. https://doi.org/10.1016/j.spmi.2018.02.018.
[11] Cho S, Kim KR, Park BG, Kang IM. RF performance and small-signal parameter extraction of junctionless silicon nanowire MOSFETs. IEEE Trans Electron Devices May 2011;58(5):1388–96. https://doi.org/10.1109/TED.2011.2109724.
[12] Rios R, Cappellani A, Armstrong M, Budrevich A, Gomez H, Pai R, Rahhal-orabi N, Kuhn K. Comparison of junctionless and conventional trigate transistors with L_{g} down to 26nm. IEEE Electron Device Lett 2011;32(9):1170–2. https://doi.org/10.1109/LED.2011.2158978. Sep.
[13] Mangal G, Tyagi A, Chaujar R. Technology computer aided design of a novel fully gate covered channel junctionless SOI FinFET for high performance analog application. In: Proceedings of the devices for integrated circuit (DevIC); May 2021. p. 406–10. https://doi.org/10.1109/DevIC50843.2021.9455801.
[14] Patterson RL, Dickman JE, Hammoud A, Gerber S. Electronic components and circuits for extreme temperature environments. In: Proceedings of the IEEE aerospace conference proceedings6; 2003. https://doi.org/10.1109/AERO.2003.1235180 *(Cat. No.03TH8652), Big Sky*p. 6_2543-6_2548.
[15] Elbuluk M, Hammoud A, Patterson R. Power electronic components, circuits and systems for deep space missions. In: *Proceedings of the* IEEE 36th conference on power electronics specialists; 2005. p. 1156–62. https://doi.org/10.1109/PESC.2005.1581775. 2005.
[16] Toan HLM, Singh SS, Maity SK. Analysis of temperature effect in quadruple gate nano-scale FinFET. Silicon 2020. https://doi.org/10.1007/s12633-020-00615-x. Aug.
[17] Kumar A, Saini S, Gupta A, Gupta N, Tripathi MM, Chaujar R. Sub-10nm high-k dielectric SOI-FinFET for high performance low power applications. In: Proceedings of the 6th international conference on signal processing and communication (ICSC); 2020. p. 310–4. https://doi.org/10.1109/ICSC48311.2020.9182748. Mar.
[18] Choi JH, Bansal A, Meterelliyoz M, Murthy J, Roy K. Self-consistent approach to leakage power and temperature estimation to predict thermal runaway in FinFET circuits. IEEE Trans Comput Aided Des Integr Circ Syst 2007;26(11):2059–68. https://doi.org/10.1109/TCAD.2007.906470. Nov.
[19] ATLAS user's manual. Santa Clara, CA, USA: SILVACO International; 2016.
[20] Madan J, Chaujar R. Gate drain-overlapped-asymmetric gate dielectric-GAA-TFET: a solution for suppressed ambipolarity and enhanced ON state behavior. Appl Phys A 2016;122(11):973. https://doi.org/10.1007/s00339-016-0510-0. Nov.
[21] Pradhan KP, Mohapatra SK, Sahu PK, Behera DK. Impact of high-k gate dielectric on analog and RF performance of nanoscale DG-MOSFET. Microelectron J 2014;45(2):144–51. https://doi.org/10.1016/j.mejo.2013.11.016. Feb.
[22] Chen Z, Wong H, Han Y, Dong S, Yang BL. Temperature dependences of threshold voltage and drain-induced barrier lowering in 60nm gate length MOS transistors. Microelectron Reliab 2014;54(6–7):1109–14. https://doi.org/10.1016/j.microrel.2013.12.005. Jun.
[23] Godoy A, López-Villanueva JA, Jiménez-Tejada JA, Palma A, Gámiz F. A simple subthreshold swing model for short channel MOSFETs. Solid State Electron 2001;45(3):391–7. https://doi.org/10.1016/S0038-1101(01)00060-0. Mar.
[24] Sreenivasulu VB, Narendar V. Design and temperature assessment of junctionless nanosheet FET for nanoscale applications. Silicon 2021. https://doi.org/10.1007/s12633-021-01145-w. May.
[25] Madan J, Chaujar R. Temperature associated reliability issues of heterogeneous gate dielectric—gate all around—tunnel FET. IEEE Trans Nanotechnol 2018;17(1):41–8. https://doi.org/10.1109/TNANO.2017.2650209. Jan.

Gaurav Mangal is currently perusing his B.Tech. from the Department of Applied Physics, Delhi Technological University. He is currently also working for the Microelectronics Research Laboratory at Delhi Technological University and his research involves modeling and simulation of nanoscale semiconductor devices for various applications.

Aman Tyagi is currently a final year undergraduate student in the Department of Applied Physics, Delhi Technological University, Delhi, India. His-current research interest is focused on nanoscale semiconductor devices for high-performance analog circuitry and sensing.

Rishu Chaujar is presently working as a Professor in the Department of Applied Physics, Delhi Technological University. Her research interests include modeling, design, and simulation of FETs, semiconductor devices, nano-electronics, etc. She has authored or co-authored more than 300 papers in various reputed international and national journals and conferences. She has supervised several National and International research projects.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Numerical study of 3-D helical passive micromixer having both inlets at offset with blood as fluid

Akash Sinha *, M. Zunaid

Department of Mechanical Engineering, Delhi Technological University, Delhi 110042, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Passive micromixer
Passive mixing
Computational fluid dynamics
Helical offset inlets channel
Blood
Simple T-micromixer (STM)
3D dimensional helical micromixer having both inlets at offset (TDHM-TO)

ABSTRACT

In the present study, two different micromixers have been evaluated for their mixing efficiencies with non-Newtonian working fluid blood. This helical micromixer has been chosen since it solves the problem of high mixing time during the mixing of fluids at the micro-level. This helical micromixer has both of its inlets at an offset to each other and is named TDHM-TO, i.e., 3D helical micromixer with inlet at offset. When this helical micromixer was compared with standard simple T-micromixer (STM), it is found that TDHM-TO requires much less mixing length than that of STM and still provides better mixing efficiency. Continuity equation, Navier-Stokes and species transport equations have been solved to investigate the mixing efficiency of the micromixers for non-Newtonian fluid blood. The study considers a wide range of mass flow rates (0.0004 kg/hr.–0.10 kg/hr.). It is being found that, for all values of mass flow rate, TDHM-TO performed much better than STM. E.g., for flow rate = 0.07 kg/hr., TDHM-TO provide a mixing efficiency of 74.41% compared to 4.93 of STM which is about 65.2% more than that of STM. STM showed a very poor result compared to TDHM-TO with only 7.64% of maximum mixing efficiency obtained with the flow rate = 0.1 kg/hr. compared to 72.84% of TDHM-TO. This TDHM-TO micromixer can be, thus, utilized in many chemicals, biochemical, and biomedical industries because of its higher efficiency and lower mixing length.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

1. Introduction

Microfluidics is the domain of science that deals with the behaviour of fluid flow inside microchannels having a length in micron, usually from 10 to 100 μm. It is an interdisciplinary field of science that includes chemistry, biology, solid mechanics, fluid mechanics etc hence it has a very wide range of applications. A micromixer finds its application in biochemical, biomedical, electronic cooling and other domains. A device called Lab-On-Chip is a very popular miniaturized device that includes actual laboratory functions such as pumping, mixing, and separation. These kinds of devices are possible mainly due to the advancement in the field of microfluidics. It has several advantages over "macro-labs" such as less use of reagent, low mixing time, small inventory requirements etc.

The micromixers are operated in a very low range of Reynolds numbers hence the flow is highly laminar. Due to the laminar flow, achieving desired mixing of fluids is a very difficult task to do. Sometimes, it may require an increased length of microchannel or more time to do the job. The process of mixing is done by molecular diffusion, which could be defined as the motion of all particles having a temperature above absolute zero through a fluid and it is a time-consuming process. This problem has gained a lot of attention from researchers and much research have been done in the particular field of science.

Although micromixers can be divided into several categories, but based on the process of mixing, they are classified into following categories- active and passive micromixers. In active micromixers, an external energy source, such as Magneto- thermodynamic, Ultrasonic, Dielectric-phoretic, etc. is applied to disturb the flow [1,2]. Their operational and fabrication cost is high and hence are not very cost-effective. Passive micromixers are stationary as they do not have any moving parts and do not require any external energy input. However, they require pressure drop to drive flow. These micromixers only use fluid pumps to generate the fluid flow. This pressure drops is generally achieved by changing the mixing interface between the fluid [2-4].

* Corresponding author.
E-mail address: sinha76akash@gmail.com (A. Sinha).

https://doi.org/10.1016/j.matpr.2022.04.431
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

Please cite this article as: A. Sinha and M. Zunaid, Numerical study of 3-D helical passive micromixer having both inlets at offset with blood as fluid, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.431

T-micromixer is probably the most common and easy type of passive micromixer and hence has attracted a lot of interest by the researchers [5-7]. It has T shaped geometry with two inlets at either end of the micromixer. Computational fluid dynamics tools have been used, mostly, to study the behaviour of fluid inside the passive micromixers. Luo et al [6] performed a numerical analysis on a T micromixer which they called "baffled T-micromixer". The baffled surface was used to provide more interfacial area to the fluid surface which in turn increased the mixing efficiency. Both mixing efficiency and pressure drop were unitized to calculate mixing ability of the fluid. Fonte et al [7] numerically studied T-micromixer with square bends for a broad range of Reynolds numbers ranging from 60 to 300 using both conventional and novel lamellae-based models. The results from both analyses were, then, analysed by comparing with the existing study. It was concluded that, although, both methods resulted in better mixing efficiency when compared with the existing study, but lamellae model was computationally more efficient. Lobasov et al. [8] did numerical analysis on a T-micromixer with different width to height aspect ratios for a wide range of Reynolds numbers, ranging from 5 to 700. It was concluded that the engulfment flow regime was observed in all the considered values of aspect ratio hence providing better efficiency than previously done studies.

There are multiple ways of increasing the mixing efficiency of a micromixer. One such way is by increasing the interfacial area between the fluid surface. This could be done by providing a curved, spiral or helical mixing channel. The curved surface of the channel imparts secondary flow to the fluid and hence the mixing efficiency increases. Hong et al. [9] numerically investigated a gourd-shape, having contracting and expanding design, micromixer for its mixing performance. The results are then compared with the existing work of T-micromixer. It was concluded that due to the vortex formation inside the micromixer, the mixing efficiency got improved for higher values of Reynolds number. This vortex was possible because of the shape of the micromixer. Yuan et al [10] used CFD to evaluate the cross-shaped micromixer. It was concluded that sufficient diffusion cannot be expected even at the high Reynolds numbers. Out of the various considered parameters, aspect ratio and blockage were found to have the most impact on the mixing efficiency.

There is another way of increasing the interfacial area between the fluids. This is done by laying down obstacles in the path of the fluid inside the micromixer. The obstacles also help in creating chaotic advection which could be defined as the phenomenon when fluid traces develop into complex fractals. Shi et al. [11] analysed the effects of different obstacles on the mixing efficiency of a Tesla shaped micromixer for a broad range of Reynolds numbers ranging from 0.2 to 200 using numerical and experimental analysis. These obstacles include diamond, rectangular and circular designs. It was concluded that sporadically positioned obstacles improve the mixing efficiency significantly. Wu et al. [12] numerically evaluated a T-micromixer with obstacles at inlets called vortex-inducing obstacles (VIOs). The effects of various geometrical parameters on the mixing were studied. It was concluded that for higher values of Reynolds number (Re > 27), the mixing efficiency increases with the increase in Reynolds number. Overall mixing got improved for all the taken values of Reynolds number when compared with the simple T-micromixer. Similar work can be found in reference [13,14].

A challenging issue that the micromixer faces is the problems associated with the fabrication as it does not use conventional manufacturing methods, rather, it requires unconventional methods of manufacturing. Tachibana et al [15] fabricated a 3D micromixer using lost-wax casting, a manufacturing method of rapid prototyping. 3D printing of hard wax resin was used for the mould preparation of lost-wax casting. By heating and washing with water, this hard was resin can be removed from a polydimethylsiloxane block. Similar techniques of micromixer fabrication can be studied in references [16,17].

From the above-mentioned literature reviews, it can be concluded that, although, several studies have been done on various types of geometries but the helically shaped micromixers still need some work. Hence in the present study, a helical micromixer has been evaluated for its mixing efficiency for a broad range of mass flow rates of blood. The results are then compared with the existing literature of T-micromixer. This helical micromixer has two inlets and the axis of these rectangular cross-section inlets are at offset. As mentioned earlier, blood has been taken as the working fluid. Both the mixing efficiency and pressure has been evaluated for the helical micromixer and has been compared with STM.

2. Micromixer geometry and specifications

Both the selected micromixers have been shown in the Fig. 1(a) and (b). Both the micromixers, STM and TDHM-TO, have a rectangular cross-sectional area of dimensions 100 × 100 μm at the inlet and 200*100 μm at the outlet. The axial length of the STM is taken to be 3000 μm (L_a).

TDHM-TO has two inlets that are at an offset to each other as can be seen from the Fig. 1(b). All geometrical parameters of TDHM-TO such as pitch (P_h), axial length (L_a) and Inlet and outlet width (W_i & W_0) has been shown in the Fig. 1(b). The diameter of the helix curve is 380 mm. For the calculation of the velocity of fluid from the Reynolds number, the hydraulic diameter of the micromixer has been taken.

2.1. Effects of stress concentration

The geometry of TDHM-TO has one drawback that is the problem of stress concentration associated with its curved structure. This stress could've been generated during the fabrication process at rounded corners and shape corners. Various analysis has been done on the generated stress concentration during the microfabrication process which can be referred to in the literature [20]. It is to be noted that the micromixer TDHM-TO will have higher stress concentration when compared with the Simple T-micromixer.

3. Numerical Modelling, validation and grid independence test

CFD tool has been used for numerical analysis to examine the behaviour of the fluid flow inside the micromixer. Following equations (from Sulekh et al. [21]) have been solved-.

(Continuity Equation)

$$\nabla.\vec{\boldsymbol{u}} = 0 \tag{1}$$

where $\boldsymbol{u}$ denotes the velocity vector of the fluid

$$\rho\left(\vec{u}\cdot\nabla\right)\vec{u} = -\nabla p + \nabla\cdot\vec{T} \tag{2}$$

Where ∇ is the del operator expressed as,

$$\nabla = \frac{\partial}{\partial x}\hat{i} + \frac{\partial}{\partial x}\hat{j} + \frac{\partial}{\partial x}\hat{k} \tag{3}$$

ρ is the density, p represents static pressure and T denotes the stress tensor which is expressed as,

$$\vec{T} = \eta\left(\nabla\vec{u} + \nabla\vec{u}^{t}\right) \tag{4}$$

where t denotes the transpose equation and η represents the viscosity of the fluid. The viscosity of the blood is shear strain rate

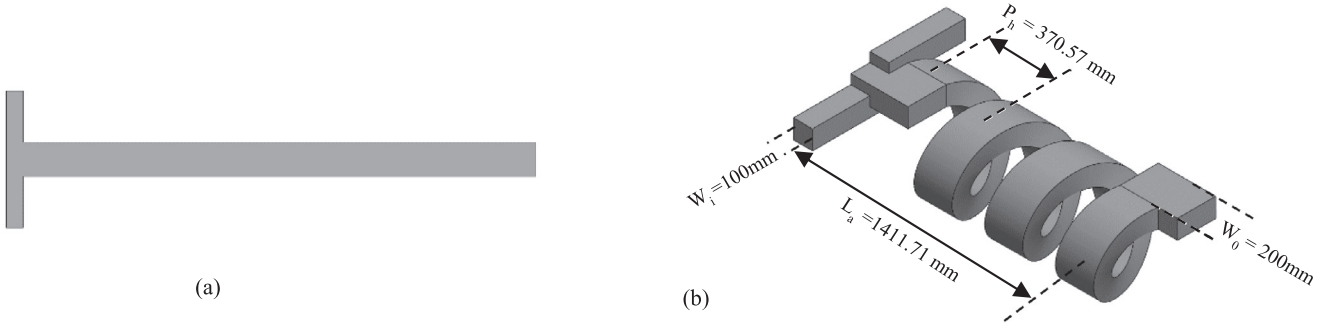


Fig. 1. Model of (a) Simple T micromixer (STM); (b) 3D Helical Micromixer having both inlets at offset (TDHM-TO) with geometrical specifications.

dependent as blood is a non-Newtonian fluid. There are several models dedicated to the study of non-Newtonian fluids. Here, the Carreau-Yasuda model has been utilized to analyze the blood flow dynamics. The literature [18] can be referred for the past research that have been done using this model. According to his model, viscosity is given by the following equation,

$$\eta(\dot{\gamma}) = \eta_0 + (\eta_0 - \eta_\infty)\left[1 + \lambda(\dot{\gamma})^a\right]^{(n-1)/a} \tag{5}$$

where η_∞ denotes infinite shear rate viscosity, η_0 denotes zero-shear viscosity, n represents the characteristic relaxation time, and a is an empirical constant. The constant values of $\eta_0, \eta_\infty, \lambda, aandn$ are 0.1600 Pa s, 0.0035 Pa s, 8.2 s, 0.64 and 0.2128 respectively. The viscosity of blood is 1,058 kg/m^3 and the value of diffusivity is taken to be 2.8 $\times$ 10^{-10} m^2/s.

Also, in the Eq. (5), $\dot{\gamma}$ is given by following equation,

$$\dot{\gamma} = \sqrt{2 \times (B : B)} \tag{6}$$

Here, B represents the tensor deformation rate. Also, the species transport model is considered in this study to study the two species mixture. The scaler transport equation of advection–diffusion is given by,

$$\left(\vec{u} \cdot \nabla\right)c = Dm\nabla^2 c \tag{7}$$

D_m represents the molecular diffusivity whereas c is the concentration.

The following equation is used to calculate the value of Mixing Index (MI) at any given equation,

$$Mi = \left(1 - \sqrt{\frac{\sigma^2}{\sigma^2_{max}}}\right) \times 100 \tag{8}$$

where

$$\sigma^2 = \frac{1}{n}\sum_{j=1}^{n}\left(c_i - c_{avg}\right)^2 \tag{9}$$

Here, σ^2 is calculated as the average sum of all the variance of the species concentration at any particular location or plane and σ^2_{max} is the maximum variance at that location. 500 points have been considered at any given location or plane. c_i and c_{avg} are the mass fraction of specie at sampling point i and average mass fraction of species concentration at that location. The value of MI signifies the degree to which the mixing has been taken place. For example, if MI is one or a hundred percent, it means that complete mixing has been taken place whereas, if it is zero, that means no mixing has been taken place. The residual for all the solved equations has been taken as 10^{-14}. To start the simulation the optimum mesh element size and mesh element numbers are needed. To calculate the mentioned parameters, a grid independence test was done. For mass flow rate 0.006 kg/hr., the simulation was performed and MI was noted for different mesh element sizes. It was observed that after three million mesh elements the change in the value of MI is very low (around 0.94%). Hence, this number of mesh elements were taken for the rest of the study. Fig. 2 shows the graph between the number of mesh elements and the value of MI. To further go with the CFD simulation, a validation of the computational method is needed to ensure that the system is running just fine. This was done by comparing the results from the present study with that of the results from the existing work. The governing equation of fluid flow and species transport were numerically solved and compared with the existing work of Afzal and Kim [19] for STM for different values of mass flow rate. The comparison can be visualized from Fig. 3. The results were found to be under satisfactory limits and hence the simulation model was validated.

4. Results and discussions

Ansys fluent solver has been used to solve various equations for a broad range of flow rates 0.00004–0.1 kg/hr. The value of mixing index, pressure drop, flow characteristics is obtained for both STM and TDHM-TO. The results of both of these micromixers are then compared with each other. As explained earlier, the Carreau-Yasuda model has been utilized to capture the behaviour of the apparent viscosity of blood. The nature of apparent viscosity concerning positive strain rate(s^{-1}) is such that it behaves as Newtonian fluid for both high (>10 s^{-1}) and low (<0.001 s^{-1}) strain rates but it shows a decreasing trend as the value of strain rate goes from low to high. There is a significant difference in the values of viscosity between water and blood and the difference is very high for lower values of strain rate.

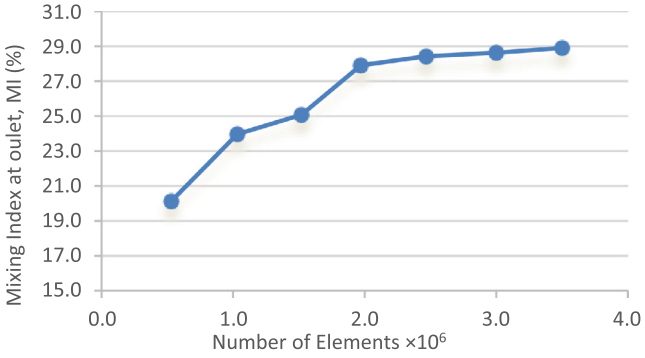


Fig. 2. Grid independence test.

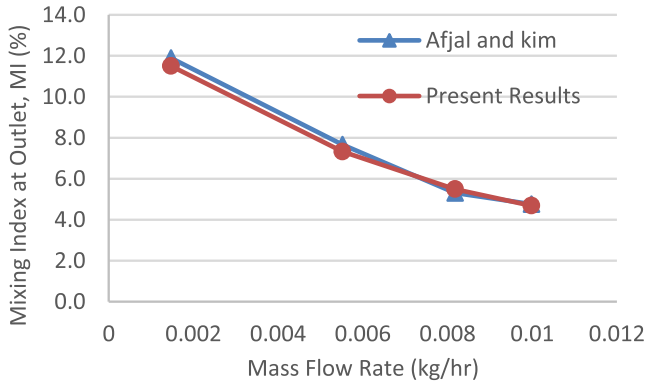


Fig. 3. Validation of numerical solution- Comparison of the current study with existing literature [19].

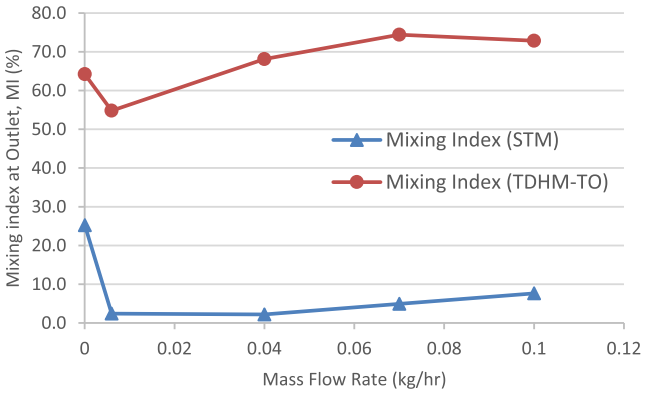


Fig. 5. Comparison of STM and TDHM-TO based on mixing index.

4.1. STM analysis with blood

Numerical analysis of STM has been done with blood as a working fluid for a range of flow rates 0.00004–0.1 kg/hr. Mixing index and pressure drop have been evaluated in the given range of values of mass flow rate. It was observed that for a low mass flow rate, the flow seemed to be parallel and straight in the mixing channel. Fig. 4(a) can be referred for visual representation for the same. The mass transport in this region is purely governed by diffusion.

The residence time is high in this region resulting in better mixing of the species. Residence time could be defined as the time spent by fluid particles inside a control volume. As the mass flow rate is increased, the residence time got decreased and hence a decrease in mixing index is observed because of the low residence time. Small vortices can be seen for mass flow rate greater than 0.04 kg/min but, as can be seen from Fig. 5, these small vortices do not have much impact on the mixing index because not much improvement was observed in MI and hence the mixing of the fluid species is still governed by the mass diffusion. The mixing index experiences a decreasing-increasing trend as the Reynolds number increases. The vortex that is being created at a higher Reynolds number due to secondary flow seems to get damped by the high viscosity of the blood and hence the increase is very low. Fig. 5 shows the graph of the variation of the mixing index Reynolds numbers.

4.2. TDHM-TO analysis with and its comparison with STM

Fig. 5 compares STM and TDHM-TO for mixing index and mass flow rate. As explained in the previous section, the advection time was high at a low Reynolds number and hence the mixing index is high for STM. The mixing was completely achieved by the mass diffusion of the molecules. The same goes for TDHM-TO: at a lower Reynolds number, the mixing efficiency was good due to the high residence time. However, as can be seen from Fig. 5, TDHM-TO performed better than STM for all values of mass flow rate. This is due to the comparatively larger interfacial area that TDHM-TO provides which helps in stretching and folding the fluid layers.

The other important reason for higher mixing efficiency is the offset inlets of TDHM-TO which creates a vortex at the inlet itself. For mass flow rate 0.07 kg/hr., the mixing efficiency of TDHM-TO was observed to be 74.41% which is greater than STM by 69.48%. This is a significant increase in the mixing efficiency. Low mixing

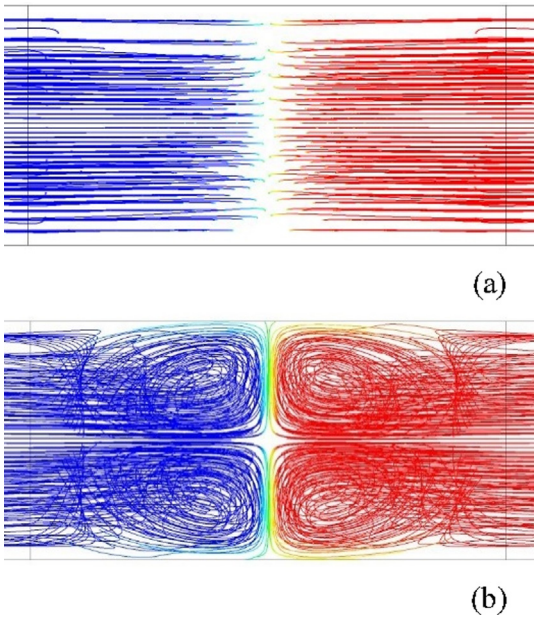


Fig. 4. Streamlines based on species concentration inside STM for mass flow rate (kg/hr.) (a) 0.0004 (b) 0.1.

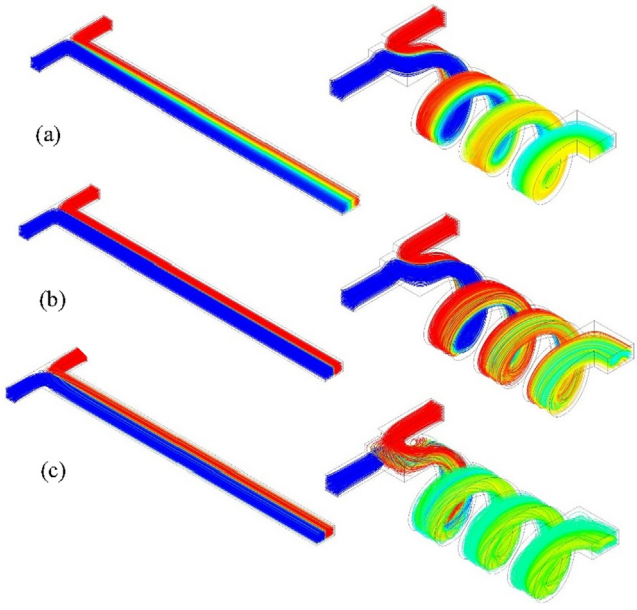


Fig. 6. Species concentration-based streamlines inside STM and TDHM-TO for mass flow rate (a) 0.00004 (b) 0.006 (c) 0.1.

efficiency was observed for the relatively higher values of Reynolds number. This is due to the low residence time caused by the high Reynolds number which caused a drop in the mass diffusion rate. Fig. 6(a), (b) & (c) show the species concentration-based streamlines inside Simple T-micromixer and TDHM-TO. As can be seen from the figure, STM experiences a decreasing-increasing trend in the mixing index as seen from Fig. 5 and the same can be observed visually from Fig. 6(a), (b) & (c) where the flowing species can be observed to be separating from each other.

TDHM-TO is experiencing a decreasing-increasing pattern where it first decreases to the lowest point and then increases due to the introduction of chaotic advection introduced because of the three-dimensional shape of the micromixer. Fig. 7(a), (b) shows the contour of species concentration at the outlet of STM and TDHM-TO and the above phenomenon can be observed from these illustrations also.

Seven different planes were selected inside the microchannel of TDHM-TO which were used to plot (Fig. 8) the species concentration to get a broad understanding of the process of mixing. The figure was plotted for a high range of mass flow rate as can be seen from Fig. 9. It can be visualized that the mixing index increases sharply from plane A to plane D for all the values of mass flow rate. For mass flow rate = 0.1 kg/hr., this increase from plane A to plane D is 47.58% which is the highest increase for any mass flow rate value. This result shows that an even smaller length microchannel can be used in some specific cases. Overall, TDHM-TO is much better compared to STM. Fig. 10(a), (b) show the contour of species concentration on the seven selected planes inside the microchannel.

When it comes to microsystems one of the most important parameters is power consumption. Power consumption in micromixer systems is directly related to the pressure drop and hence pressure drop becomes an important parameter to analyse. The increased mixing required more pumping power to drive the same amount of fluid. Micromotors are used to pump the fluids. Pressure drop provides an idea of the trend of power consumption, whether it will increase or decrease. Pressure drops inside Simple T- micromixer and TDHM-TO have been shown in Fig. 11. It was observed that no noticeable difference was seen for low mass flow rate whereas the difference in pressure drop increases in both the micromixer as the flow rate increases. It can be observed from Fig. 11 that TDHM-TO showed a relatively higher pressure drop for the values of mass flow rate.

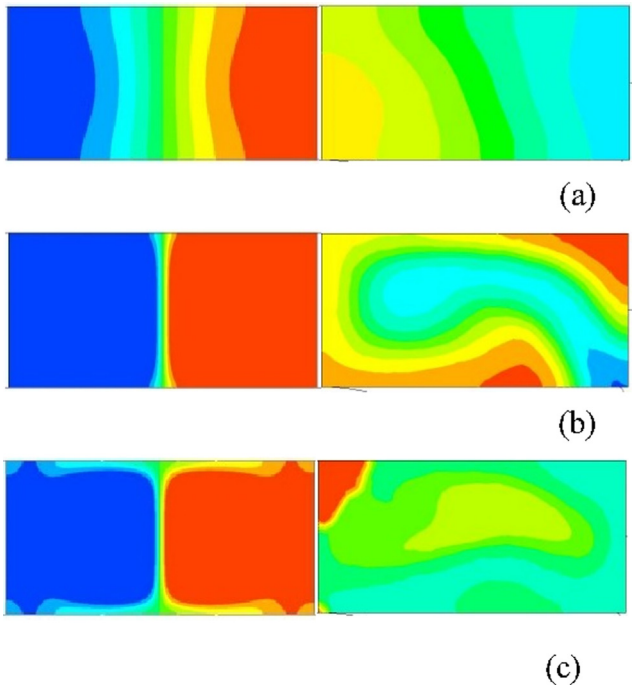


Fig. 7. Species concentration-based contour at outlet of STM and TDHM-TO for mass flow rate (a) 0.00004 (b) 0.006 (c) 0.1.

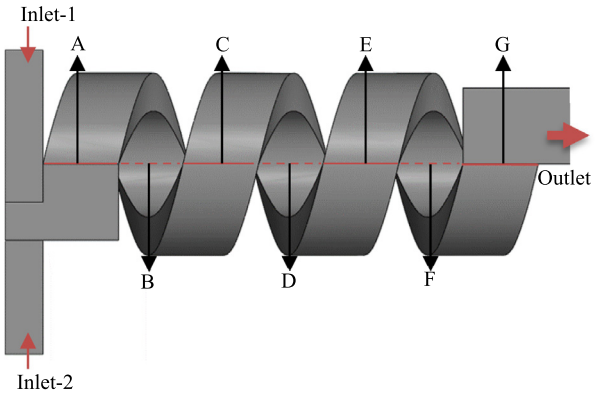


Fig. 8. Planes at different sections (Top View of TDHM-TO).

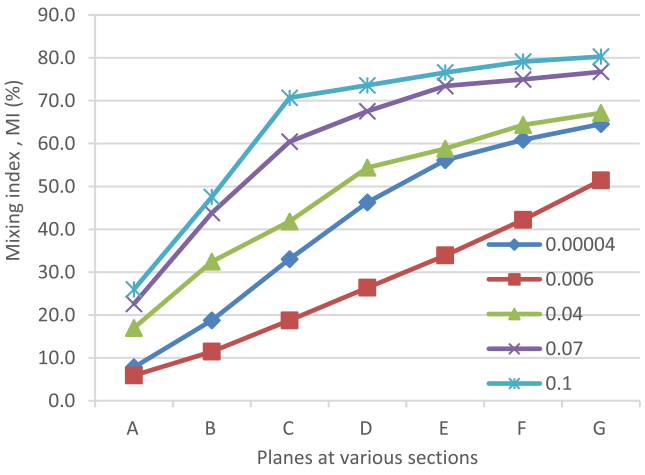


Fig. 9. Mixing Index at seven different planes inside TDHM-TO.

5. Conclusions

In this paper two micromixers, STM and TDHM-TO, have been studied and compared for their mixing efficiencies. The study takes blood as a working fluid. Following conclusions can be drawn from the results. 3-D Helical Passive Micromixer having Both Inlets at Offset (TDHM-TO) provides better efficiency for all values of mass flow rate. This is due to the offset inlet design, the larger interfacial area provided by the three-dimensional shape of the micromixer and due to the presence of secondary flow that created chaotic advection. 69.48% of significant improvement was observed for mass flow rate = 0.07 kg/hr. An increase of 47.58% from plane A to plane D was observed for mass flow rate = 0.1 kg/hr. which shows that a smaller length microchannel can also be used. The results also conclude that TDHM-TO requires a smaller axial length compared to STM and still produces better mixing results.

CRediT authorship contribution statement

Akash Sinha: Data curation, Writing – original draft, Visualization, Investigation, Software, Validation. **M. Zunaid:** Conceptual-

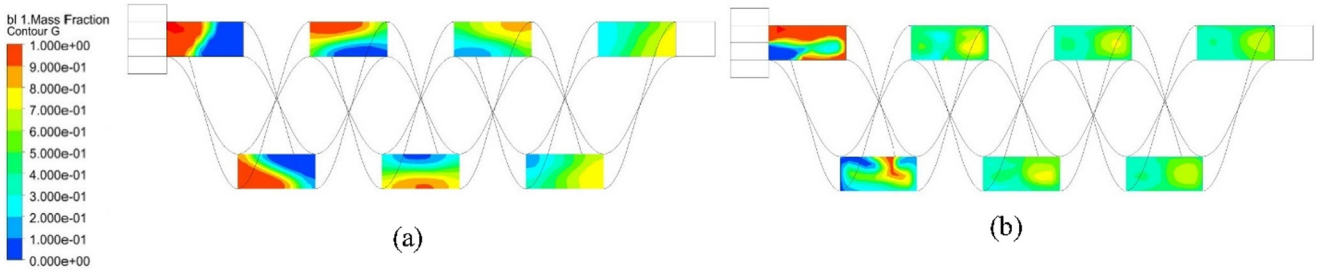


Fig. 10. Contour of species concentration at different planes A to G inside TDHM-TO for mass flow rate = (a) 0.0004 (b) 0.1.

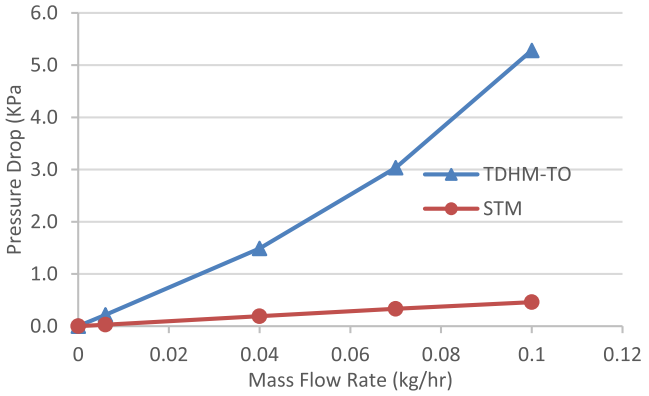


Fig. 11. Comparison of pressure drop between STM and TDHM-TO for a range of mass flow rate.

ization, Methodology, Software, Supervision, Writing – review & editing.

Declaration of Competing Interest

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

Acknowledgement

We gratefully thank Delhi Technological University for their support.

References

[1] M. Annabestani, H. Mohammadzadeh, A. Aghassizadeh, S. Azizmohseni, M. Fardmanesh, Active microfluidic micromixer design using ionic polymer-metal composites, 27th Iranian Conference on Electrical Engineering (ICEE) (2019) 371–375.
[2] I. Shah, H.S. Jeon, M. Ali, D.H. Yang, K.-H. Choi, Optimal parametric mixing analysis of active and passive micromixers using Taguchi method, Proc. Inst. Mech. Eng., Part E: J. Process Mech. Eng. 233 (6) (2019) 1292–1303.
[3] W. Raza, S. Hossain, K.-Y. Kim, A review of passive micromixers with a comparative analysis, Micromachines 11 (5) (2020) 455.
[4] W. Ruijin, L. Beiqi, S. Dongdong, Z. Zefei, Investigation on the splitting-merging passive micromixer based on Baker's transformation, Sens. Actuators, B 249 (2017) 395–404.
[5] V. Goodarzi, S.H. Jafarbeygi, R.A. Taheri, M. Sheremet, M. Ghalambaz, Numerical investigation of mixing by induced electrokinetic flow in T-micromixer with conductive curved arc plate, Symmetry 13 (2021) 915.
[6] X. Luo, Y. Cheng, W. Zhang, K. Li, P. Wang, W. Zhao, Mixing performance analysis of the novel passive micromixer designed by applying fuzzy grey relational analysis, Int. J. Heat Mass Transf. 178 (2021) 121638.
[7] C.P. Fonte, D.F. Fletcher, P. Guichardon, J. Aubin, Simulation of micromixing in a T-mixer under laminar flow conditions, Chem. Eng. Sci. 222 (2020) 115706.
[8] A.S. Lobasov, A.V. Minakov, V.V. Kuznetsov, V.Y. Rudyak, A.A. Shebeleva, Investigation of mixing efficiency and pressure drop in T-shaped micromixers, Chem. Eng. Process. - Process Intensif. 134 (2018) 105–114.
[9] W. Hong, H. Shi, Z. Huang, LihongYu, M. Long, H. Xu, Z. Liu, Design and simulation of a passive micromixer with gourd-shaped channel, J. Nanosci. Nanotechnol. 19 (1) (2019) 206–212.
[10] S. Yuan, B. Jiang, T. Peng, Q. Li, M. Zhou, An investigation of flow patterns and mixing characteristics in a cross-shaped micromixer within the laminar regime, Micromachines 12 (4) (2021) 462.
[11] X. Shi, S. Huang, L. Wang, F. Li, Numerical analysis of passive micromixer with novel obstacle design, J. Dispersion Sci. Technol. 42 (3) (2021) 440–456.
[12] C.Y. Wu, B.H. Lai, Numerical study of T-shaped micromixers with vortex-inducing obstacles in the inlet channels, Micromachines 11 (12) (2020) 1122.
[13] Z. Wu, X. Chen, A novel design for 3D passive micromixer based on Cantor fractal structure, Microsyst. Technol. 25 (1) (2019) 225–236.
[14] S. Baheri Islami, M. Khezerloo, Enhancement of mixing performance of non-newtonian fluids using curving and grooving of microchannels, J. Appl. Fluid Mech. 10 (1) (2017) 127–141.
[15] D. Tachibana, K. Matsubara, R. Matsuda, T. Furukawa, S. Maruo, Y. Tanaka, O. Fuchiwaki, H. Ota, 3D helical micromixer fabricated by micro lost-wax casting, Adv. Mater. Technol. 5 (1) (2020) 1900794.
[16] W. Guo, L. Tang, B. Zhou, Y. Fung, Fundamental studies of rapidly fabricated on-chip passive micromixer for modular microfluidics, Micromachines 12 (2021) 153.
[17] M. Rafeie, M. Welleweerd, A. Hassanzadeh-Barforoushi, M. Asadnia, W. Olthuis, M. Ebrahimi Warkiani, An easily fabricated three-dimensional threaded lemniscate-shaped micromixer for a wide range of flow rates, Biomicrofluidics 11 (1) (2017) 014108.
[18] F. Abraham, M. Behr, M. Heinkenschloss, Shape optimization in steady blood flow: A numerical study of non-Newtonian effects, Comput. Methods Biomech. Biomed. Eng. 8 (2) (2005) 127–137.
[19] A. Afzal, K.Y. Kim, Flow and mixing analysis of non-Newtonian fluids in straight and serpentine microchannels, Chem. Eng. Sci. 116 (2014) 263–274.
[20] M. Kumar, J. Rajiyan, P. Gupta, A computational approach for solving elasto-statics problems, Mater. Today:. Proc. 46 (15) (2021) 6876–6879.
[21] S. Tokas, M. Zunaid, M.A. Ansari, Numerical investigation of the performance of 3D-helical passive micromixer with Newtonian fluid and non-Newtonian fluid blood, Asia-Pac. J. Chem. Eng. 16 (1) (2021).

Silicon
https://doi.org/10.1007/s12633-022-01930-1

ORIGINAL PAPER

Performance Evaluation & Linearity Distortion Analysis for Plasma-Assisted Dual-Material Carbon Nanotube Field Effect Transistor with a SiO_2-HfO_2 Stacked Gate-Oxide Structure (DM-SGCNFET)

Mansha Kansal[1] · Suresh C. Sharma[1]

Received: 27 January 2022 / Accepted: 6 May 2022
© The Author(s), under exclusive licence to Springer Nature B.V. 2022

Abstract
This work demonstrates the simulation analysis of a novel device Plasma-Assisted Dual-Material Stacked Gate-Oxide Carbon Nanotube Field Effect Transistor (DM-SGCNFET) with a stacked gate oxide structure consisting of two oxides - SiO_2 and HfO_2. The performance metric of the simulated device is compared with a SiO_2 based similar device (DM-CNFET) for varying sets of plasma parameters that corresponds to different values of channel length. Both the devices feature a PECVD grown vertically aligned semiconducting carbon nanotube implemented as the channel material. The comprehensive analysis delineates that the DM-SGCNFET offers enhanced values of drain current, transconductance, output conductance, early voltage, gain, gate capacitance, better switching ratio, and reduced channel resistance as compared to DM-CNFET. The novel structure contemplating a dual-material gate and a high-k dielectric gate stack in CNT based transistors warrants as an appropriate device for digital and analog applications providing high efficiency. The paper further explores the DM-SGCNFET device for linearity distortion performance by evaluating g_{m2}, g_{m3}, VIP2, VIP3, IIP3, IMD3 and 1-dB compression point for altering values of plasma parameters.

Keywords Plasma parameters · Gate stack · Vertically aligned carbon nanotube · Carbon nanotube field effect transistor (CNTFET) · Plasma enhanced chemical vapor deposition (PECVD)

1 Introduction

Carbon Nanotubes (CNTs) have emerged as a lucrative option to the ever growing global trend of downscaling the physical dimensions of transistors while simultaneously aiming for increased device performance [1]. Owing to their capable electrical and mechanical properties, the rise of CNTs has made nanotechnology more viable, thus opening new avenues of real life applications as well. CNT based devices offer a host of advantages making them popular for nanotechnology and nanoarchitectonics. They showcase high speed operation, high mechanical flexibility, high current density, excellent thermal conductivity etc. [2, 3]. The field of nanoarchitectonics is an emerging concept which uses the concept of nanotechnology for the production of systems built from nanoscale objects. In simpler words, nanoarchitechtonics is used to create real life practical systems from nanotechnology. After years of extensive research on silicon and silicon-based devices, CNT based devices have gathered massive interest and emerged as viable alternatives [4, 5].

CNTs can be synthesized using a variety of ways. Among these, PECVD (Plasma Enhanced Chemical Vapor Deposition) technique stands out owing to a variety of reasons. This particular synthesis technique allows for large scale production of vertically aligned CNT structures with relatively better control over the structure's dimensions and physical properties, that too at a lower temperature. This is made possible owing to the existence of diverse combination of some active species - electrons, ions, radicals, neutrals etc. within the plasma region. The electric field of the plasma sheath is responsible for applying force on the carbon nanostructures, which leads to their growth in a vertical direction [6–12].

✉ Suresh C. Sharma
prof_sureshsharma@dtu.ac.in

Mansha Kansal
kansal.mansha@gmail.com

[1] Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110085, India

Published online: 21 May 2022

The robust nature of SiO_2 has allowed for its extensive usage in CMOS technology. In spite of the significant advantages it holds, it suffers from a relatively lower dielectric constant. Since capacitance is inversely proportional to dielectric thickness, SiO_2 needs to be scaled down accordingly. SiO_2 however, suffers from a variety of scaling problems such as increasing gate leakage current, increasing direct tunnelling current, reliability degradation etc. This can lead to erroneous operation of logic device and an increase in static power consumption [13].

Innovations through novel materials and structures are paving the way for significant advancements in performance and reliability. One such innovation is the usage of high-k dielectrics as gate oxide. Out of the various viable high-k dielectrics, HfO_2 has emerged as a promising material. It possesses some remarkable properties such as high dielectric constant ($\kappa \sim 22$), high thermal stability, comparably large bandgap (5.3 eV), and also high heat of formation [14].

While high-k materials such as HfO_2 has its merits, on the flip side, its direct deposition is not so favourable. A head-on deposition can suffer from unstable threshold voltage, degradation in mobility and subpar subthreshold swing. Moreover, an increased thickness of high-k dielectrics leads to large magnitudes of fringing fields, thus causing poor gate controllability [15]. Thus, implementing a stacked gate oxide structure is preferred. A gate stack structure requires a thin SiO_2 layer between the high-k dielectric and the channel material, thus taking advantage of superior SiO_2 interface quality.

In this paper, Plasma-Assisted Dual-Material Stacked Gate-Oxide Carbon Nanotube Field Effect Transistor (DM-SGCNFET) has been simulated using SILVACO ATLAS 3D simulator. The device features a cylindrical surrounding gate structure. Further shown is the variation of electrical properties of device such as drain current, channel resistance, I_{ON}/I_{OFF} ratio, transconductance, output conductance, early voltage, gain and gate capacitance with plasma parameters - electron density n_{oe}, electron temperature T_e and ion density n_{oi}, ion temperature T_i, corresponding to different sets of channel length at fixed channel radius. The results have been compared to SiO_2 based Plasma-Assisted Dual-Material Carbon Nanotube Field Effect Transistor (DM-CNFET). It can be seen that due to employment of stacked gate oxide structure, there is considerable improvement in performance over DM-CNFET. This work further examines the effect of plasma parameters on the wireless operation of the proposed device using linearity and harmonic distortion parameters - g_{m2}, g_{m3}, VIP2, VIP3, IIP3, IMD3 and 1-dB compression point.

2 Device Simulation Framework

The proposed device DM-SGCNFET is designed and simulated. The experimental synthesis of carbon nanotubes has been done using PECVD technique at relatively low temperature using Argon and Methane as precursor gases on a silicon substrate with nickel deposited by thermal evaporation [16]. The 3-D and cross-sectional view of DM-SGCNFET are depicted in Figs. 1 & 2, respectively.

The contour plot of potential at $V_{gs} = 0$ V and $V_{ds} = 0$ V is shown in Fig. 3. (a) whereas the contour plot for hole concentration at $V_{gs} = -1.0$ V and $V_{ds} = 1.0$ V is shown in Fig. 3. (b). The proposed device uses a dual-material gate having work functions 4.4 eV and 4.7 eV, with the former at the drain side and latter at the source end. Table 1 shows the structure specifications of the device DM-SGCNFET. SiO_2 ($\kappa = 3.9$) and HfO_2 ($\kappa = 22$) are used to simulate a stacked gate oxide structure. HfO_2 has been deployed above SiO_2 in order to increase the oxide capacitance of the device.

The source and drain contacts are used with doping concentration $10^{18}/cm^3$. Silvaco Atlas TCAD software is used to run the device simulations [17]. In order to successfully simulate the device and get accurate results, defining a mesh of the structure becomes critical. It defines how dense the grid will be at the time of calculation. A smaller mesh ensures a dense grid, thus resulting in better accuracy at the cost of increased time for completion. It is a balance of accuracy and time that decides the perfect mesh structure. By specifying a uniform doping profile in the simulation software, the source and drain contacts are doped with n type impurity. By using MUP and MUN parameters, mobility for holes and electrons is defined for simulating CNT. The leakage currents are simulated using Shockley-Read-Hall (SRH) model while minority carrier recombinations are factored in using the Auger recombination model. The Boltzmann model is implemented and Newton and Gummel's method are incorporated for mathematical solutions [18]. The models used in this manuscript are tabulated in Table 2.

The results obtained after simulating the device DM-SGCNFET, as shown in Fig. 4, can be seen following the

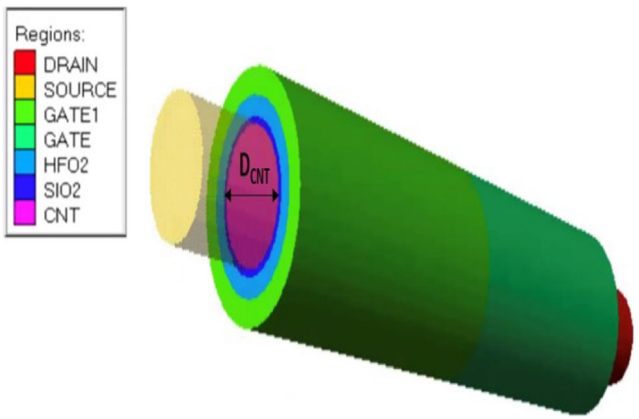


Fig. 1 3-D view of the proposed DM-SGCNFET

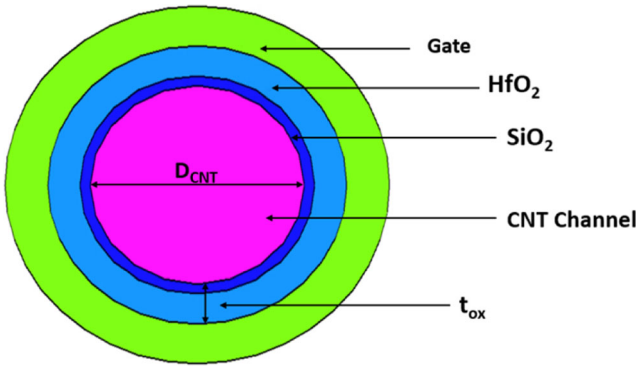


Fig. 2 Depiction of cross sectional view of DM-SGCNFET

same trend as the device reported by Yang et al. and Kansal et al. [18, 19]. The difference in magnitude is owing to basic structural differences that exist between the two devices.

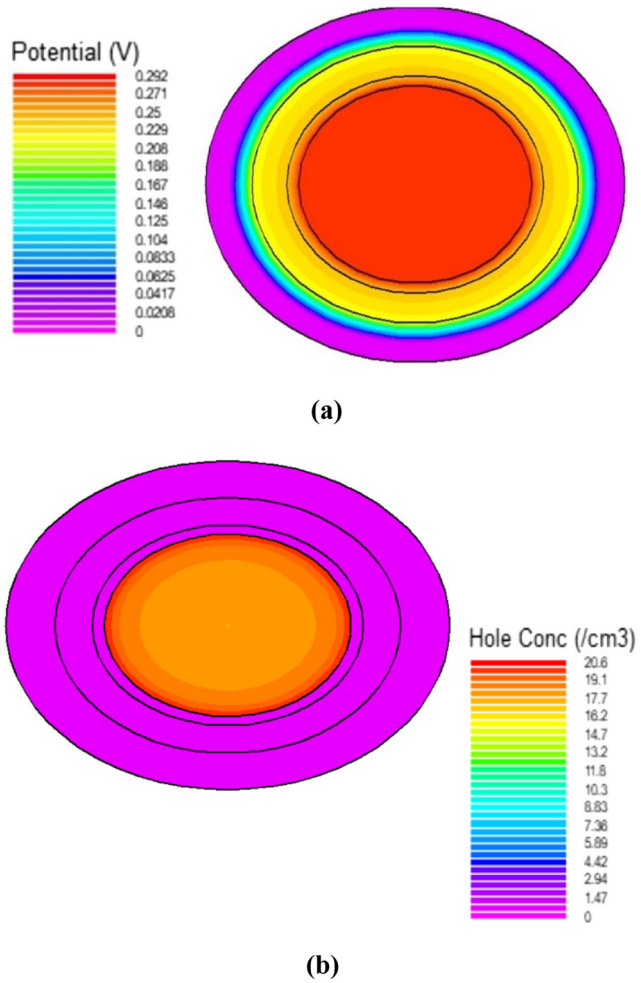


Fig. 3 **a** Potential contour plot of DM-SGCNFET at V_{ds} = 0 V and V_{gs} = 0 V **b** Contour plot for hole concentration of DM-SGCNFET at V_{gs} = −1.0 V and V_{ds} = 1.0 V

Table 1 Device Specifications

Parameter	Symbol	Value(s)
Channel Length (μm)	L	1.5, 2.5, 3.5
Channel Radius (nm)	R	5
Oxide thickness (nm)	t_{ox}	2
Source/ Drain Length (nm)	L_s/L_d	15
Gate work function (eV)	ϕ_1, ϕ_2	4.4, 4.7
SiO_2 Permittivity	ε_{ox1}	3.9
HfO_2 Permittivity	ε_{ox2}	22

3 Results and Discussion

The structure and characteristics of the proposed device DM-SGCNFET are simulated and analysed using the Silvaco 3D Atlas TCAD Simulator.

The first part of the research aims to analyse the characteristics of SiO_2 based DM-CNFET with varying oxide thickness. From Fig. 5. (a-c), it is observed that there is an improvement in the performance of the device on reducing the oxide thickness. However, it is known that on reducing the oxide thickness, even for improved performance, gate tunneling current & gate leakage current increases. Thus, it becomes difficult to reduce or scale down the dimensions of the device without risking its breakdown caused by the inability to further reduce the oxide thickness. Moreover, SiO_2 is predicted to reach the limits of its scaling after which it will act as a hindrance in the continued downscaling trend. Owing to these limitations, research is now focused more towards the advent of new material technologies than continuing to usher in the era of existing ones. High-k oxide materials, as a result, have sparked tremendous amounts of interest [20].

With the help of a physical model as developed by Sharma and Tiwari [21], the relation between plasma parameters and device geometry can be drawn. With increment in values of plasma parameters: electron density n_{oe} and electron temperature T_e, ion density n_{oi} and ion temperature T_i, the CNT length increases initially with time before hitting the saturation level. For increasing plasma parameters: n_{oi} and T_i, there is an increase in the thickness of the plasma sheath and there is a faster generation of carbon monomers. Consequently, there is more diffusion of the carbon ions taking place, hence leading to increase in length of the CNT [10, 22].

Plasma parameters play a severely important role in deciding the performance of the synthesized device. Owing to their control over the channel radius and channel length, they have a direct control over critical electrostatic properties such as gate controllability, contact resistance, conductivity etc. The intention is to further cement our knowledge about the impact of plasma parameters by analyzing the characteristics of DM-SGCNFET over varying sets of plasma parameters

Table 2 Simulation Models required for simulations

Simulation Model	Description
SRH (Schottky-Read-Hall) Model	Appropriate for taking account of recombination effects and simulates the leakage currents
Auger-Model	The minority carrier recombinations are factored in using this model
Boltzmann-Model	Appropriate to consider carrier statistics
Mobility Model	MUN and MUP parameters are used to define electron and hole mobility

corresponding to various values of channel length, at a fixed channel radius.

Following Kansal and Sharma [18], surface potential can be defined as:

$$V_s = -\frac{2e}{L}\log\left(\frac{\lambda_D}{R}\right) \tag{1}$$

e = electronic charge, λ_D= Debye length, L = channel length, R = channel radius.

From Eq. (1),

$$\mathrm{L} = -\frac{2e}{V_s}\log\left(\frac{\lambda_D}{R}\right), \tag{2}$$

Also, by the definition of Debye length:

$$\lambda_D^{-2} = \lambda_{de}^{-2} + \lambda_{di}^{-2} \tag{3}$$

where λ_{de}= electron plasma Debye length,
λ_{di}= ion plasma Debye length

$$\lambda_{de} = \sqrt{\frac{T_e}{4\pi n_{oe}e^2}} \;\&\; \lambda_{di} = \sqrt{\frac{T_i}{4\pi n_{oi}e^2}} \tag{4}$$

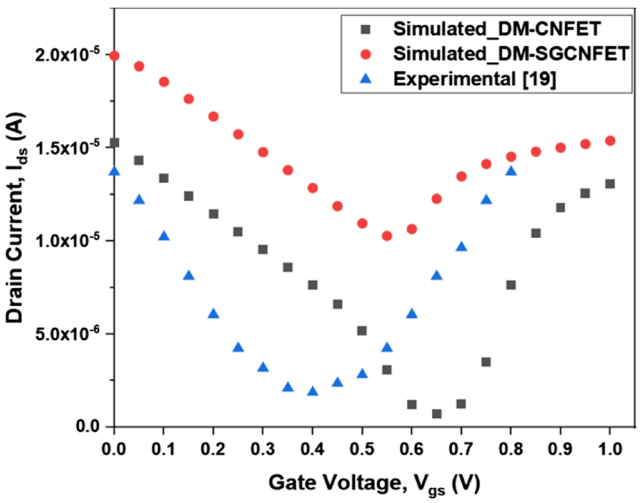


Fig. 4 Comparison of simulated devices with experimental results at V_{ds} = 0.8 V

By using Eq. (3), Eq. (2) becomes:

$$L = -\frac{2e}{V_s}\log\left[\frac{1}{R}\left(\frac{\lambda_{de}\ \lambda_{di}}{\sqrt{\lambda_{de}^2 + \lambda_{di}^2}}\right)\right] \tag{5}$$

$$L = -\frac{e}{V_s}\log\left[\frac{1}{R^2}\left(\frac{\lambda_{de}^2\ \lambda_{di}^2}{\lambda_{de}^2 + \lambda_{di}^2}\right)\right]$$

$$L = \frac{e}{V_s}\log\left[R^2\left(\frac{\lambda_{de}^2 + \lambda_{di}^2}{\lambda_{de}^2\ \lambda_{di}^2}\right)\right]$$

$$L = \frac{e}{V_s}\log\left[R^2\left(\frac{1}{\lambda_{de}^2} + \frac{1}{\lambda_{di}^2}\right)\right] \tag{6}$$

Putting the values of λ_{de} and λ_{di} from Eq. (4), Eq.(6) becomes as follows:

$$L = \frac{e}{V_s}\log\left[R^2\left(\frac{4\pi n_{oe}e^2}{T_e} + \frac{4\pi n_{oi}e^2}{T_i}\right)\right]$$
$$L = \frac{e}{V_s}\log\left[4\pi e^2R^2\left(\frac{n_{oe}}{T_e} + \frac{n_{oi}}{T_i}\right)\right] \tag{7}$$

The drain current is formulated as:

$$I_{ds} = \frac{\mu C}{L^2}V_{ds}\left(V_{gs} - V_{th}\right), \tag{8}$$

C = Capacitance of the CNT, μ = apparent mobility in linear regime and V_{th} = threshold voltage.

By putting the value of channel length L from Eq. (7) in Eq. (8), it becomes possible to relate the drain current of the device with the plasma parameters. As other characteristics of any device such as transconductance (g_m), output conductance (g_d), early voltage (V_{EA}) etc. depend on the drain current, so we can further draw a relation between these transistor properties and the plasma parameters.

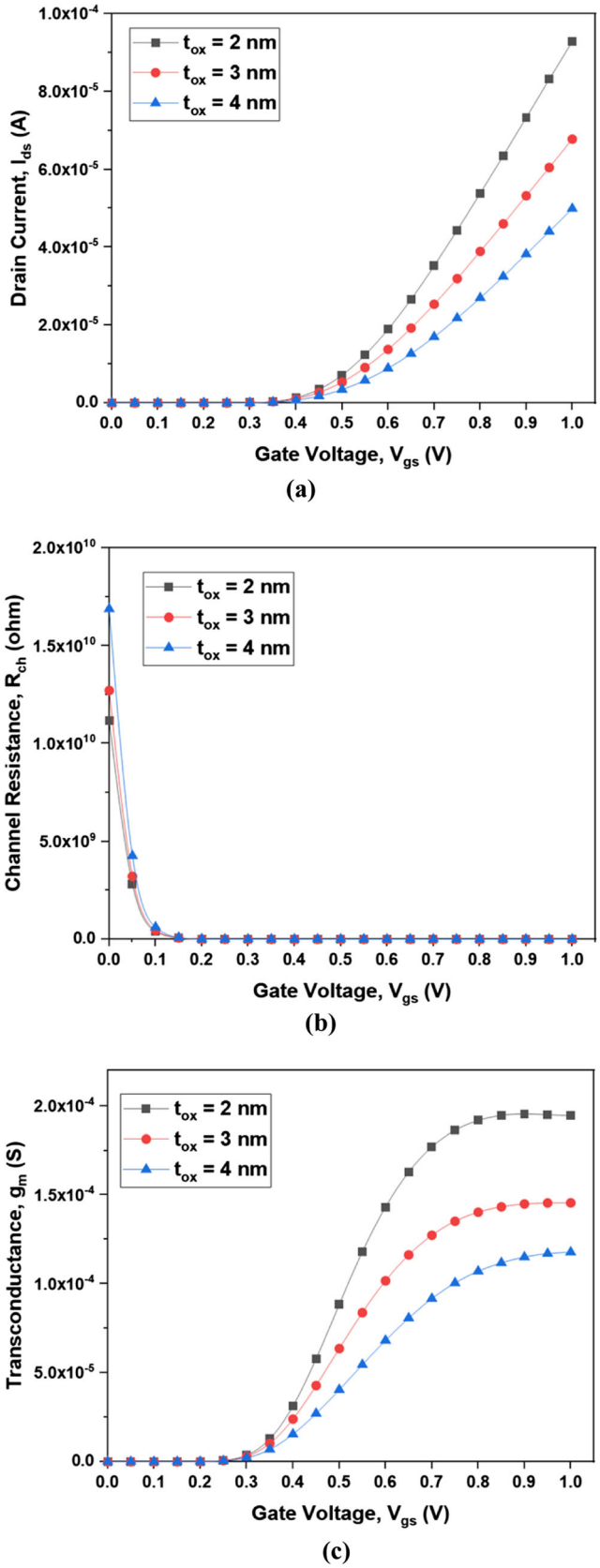


Fig. 5 Impact of oxide thickness variation on SiO_2 based DM-CNFET: (**a**) Drain current I_{ds} v/s gate voltage V_{gs} (**b**) Channel Resistance R_{ch} v/s V_{gs} (**c**) Transconductance g_m v/s V_{gs} (at constant V_{ds} = 0.8 V)

The transfer and output characteristics for the DM-SGCNFET device at V_{ds} = 0.8 V and V_{gs} = 0.5 V, respectively for varying sets of plasma parameters is depicted in Fig. 6. (a) and Fig. 6. (b), respectively. From Fig. 6. (c), it is observed that the result of analytical expression obtained for drain current in linear regime mentioned in Eq. 8 validates the simulation result for Case 1 shown in Table 3. It is further revealed that the proposed device shows considerable enhancement in drain current and degradation in channel resistance as can be seen from Fig. 7 when compared to SiO_2 based DM-CNFET device. This is due to the incorporation of gate stack scheme having high-k dielectric. HfO_2 has band gap of 5.3 eV whereas the band gap for SiO_2 is 9.0 eV [23]. Due to lower energy gap in DM-SGCNFET, there is a lower energy barrier than DM-CNFET, which leads to faster carrier generation and hence increased drain current can be encountered in the simulated device.

Figure 8 exhibits the I_{ON}/I_{OFF} ratio for both device structures. Low values of I_{ON}/I_{OFF} ratio causes low output transitions, whereas an appropriate value can help in improving speed and minimize leakage. It can be defined as the ratio of ON state current (value of I_{ds} at V_{gs} = 1.0 V) that determines the current driving capability of the device and OFF state current (value of I_{ds} at V_{gs} = 0 V) that refers to the power dissipation. Though there is a slight increase in the leakage current on using the SiO_2-HfO_2 stacked gate oxide but the increase in I_{ON} is much higher for the proposed device, hence leading to improved values of the switching ratio.

Transconductance (g_m) is a crucial parameter used to define the optimum bias point. It is also an important measure for RF and analog applications. Figure 9 outlines the plot of variation of g_m with V_{gs} for different sets of plasma parameters at V_{ds} = 0.8 V for SiO_2 based DM-CNFET and DM-SGCNFET devices. g_m tends to follow an increasing trend with drop in the values of plasma parameters and it can be observed that DM-SGCNFET shows improved values of g_m in comparison to DM-CNFET due to high values of drain current as reported for former device. The increasing values of I_{ds} for the simulated device is attributed to the higher capacitance of HfO_2 as it leads to a hike in the electron concentration that takes place in the channel.

Output conductance, g_d is another parameter that determines the driving ability of a device and can be denoted as the derivative of drain current to drain bias at fixed V_{gs}. As g_d is directly dependant on drain current, any fluctuation in drain current due to employment of high-k dielectric stack would consequently lead to change in output conductance. In Fig. 10, the value of g_d increases with scaling down of plasma parameters and the simulated device shows enhancement in g_d due to incorporation of high-k gate stack. Owing to high permittivity of the oxide layer, there is an increase in capacitance leading to creation of higher electric field, which further leads

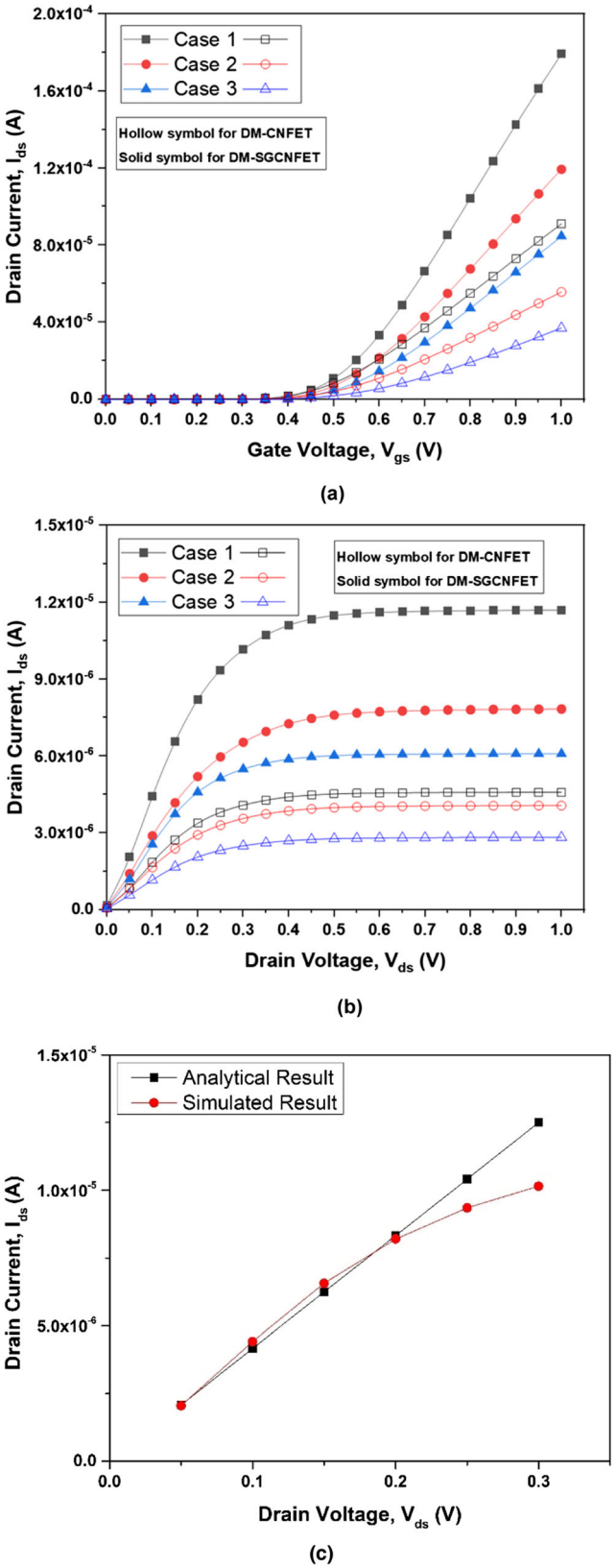


Fig. 6 Impact of plasma parameters variation (Table 3) on (**a**) transfer characteristics at V_{ds} = 0.8 V (**b**) output characteristics at V_{gs} = 0.5 V (**c**) Comparison of simulated and analytical results in linear regime at V_{gs} = 0.5 V for Case 1

to enhancement in the electron velocity. Hence, the electrons traverse faster and consequently increasing the I_{ds}.

Early voltage V_{EA} plays a significant role in determining the analog performance of a device. It is formulated as the ratio of drain current I_{ds} and output conductance g_d [24]. Higher V_{EA} implies enhanced gain of the device and its value improves with increasing set of plasma parameters. This can be attributed to the fact that g_d decreases at a more pronounced rate as compared to I_{ds}. From Fig. 11, it is observed that DM-SGCNFET exhibits better values of early voltage in comparison to DM-CNFET making it an appropriate device for OPAMP applications.

The variation of the total Gate Capacitance (C_{GG}) that comprises of both gate to source capacitance (C_{GS}) and gate to drain capacitance (C_{GD}) with V_{gs} for both the devices for different set of plasma parameters is depicted in Fig. 12. The simulation using AC analysis of C_{GG} is executed by coupling an input small AC signal with DC bias at the gate terminal. It can be clearly observed from the figure that the simulated device DM-SGCNFET offers large C_{GG} due to its higher gate capacitance as the fringing field capacitances comes into picture in addition to the above mentioned capacitances C_{GS} and C_{GD} on employment of high-k gate stack structure [25].

3.1 Impact of Plasma Parameters on Linearity Characteristics of DM-SGCNFET

Transistors with low distortion parameters are quite desirable for wireless communication systems. The parameters like high-order derivatives of transconductance (g_{m2} and g_{m3}) exhibit non-linearity and hence amplitude of these should be minimal in order to safeguard linearity and for lesser distortion. The nonlinearity effects lead to the degradation of the device characteristic and performance. Fig. 13(a) and (b) show the effect of plasma parameters on g_{m2} and g_{m3} with respect to V_{gs}. The values of g_{m2} and g_{m3} decrease with increasing sets of plasma parameters. At higher gate voltage, g_{m3} attains negative peak values.

The Figure of Merit (FOM) VIP2, the second-order voltage intercept point and VIP3 known as the third-order voltage intercept point can be defined as the extrapolated gate voltage at which the fundamental tone amplitude is equal to the second-order and third-order harmonic amplitude respectively and can be used for the assessment of the characteristics of signal distortion using DC parameters [26]. The high peak values of these parameters VIP2 and VIP3 indicate better linearity characteristics of the device.

The device FOM can be formulated as [26, 27]:

$$VIP2 = 4 \times \frac{g_m}{g_{m2}} \tag{9}$$

$$VIP3 = \sqrt{24 \times (g_m/g_{m3})} \tag{10}$$

Table 3 Value of Plasma parameters corresponding to varying values of channel length (at constant value of R = 5 nm)

Scenario	Channel Length (μm)	Plasma Parameter I - electron density n_{oe} (cm^{-3})	Plasma Parameter II - electron temperature T_e (K)	Plasma Parameter III - ion density n_{oi} (cm^{-3})	Plasma Parameter IV - ion temperature T_i (K)
Case 1	1.5	8×10^{8}	1.33×10^{4}	7×10^{7}	2200
Case 2	2.5	2×10^{9}	1.45×10^{4}	6×10^{8}	2235
Case 3	3.5	8×10^{9}	1.56×10^{4}	2×10^{9}	2280

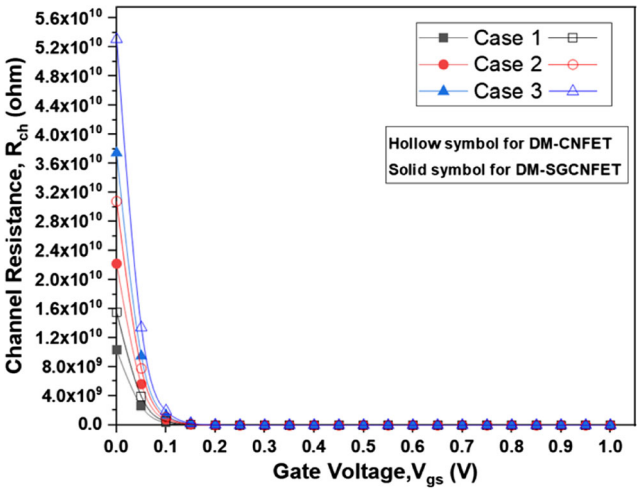


Fig. 7 Impact of different values of plasma parameters (Table 3) on channel resistance (R_{ch}) at V_{ds} = 0.8 V

$$IIP3 = \frac{2}{3} \times \frac{g_m}{g_{m3} \times R_s} \tag{11}$$

$$IMD3 = \left[(9/2) \times (VIP3)^2 \times g_{m3}\right]^2 \times R_s \tag{12}$$

$$1\text{−dB compression point} = 0.22 \times \sqrt{(g_m/g_{m3})} \tag{13}$$

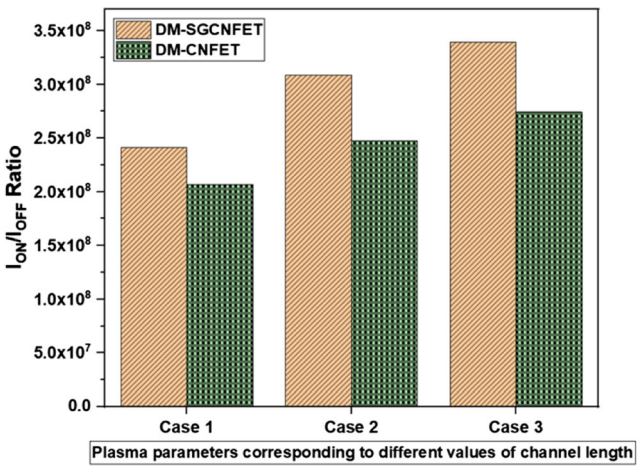


Fig. 8 Variation of I_{ON}/I_{OFF} Ratio for different values of plasma parameters (Table 3) at V_{ds} = 0.8 V

where R_s is known as the internal resistance and its value can be taken as 50 Ω.

In Fig. 13(c) and (d), VIP2 and VIP3 variation with respect to V_{gs} at V_{ds} = 0.8 V has been plotted for different sets of plasma parameters, respectively. A higher value of VIP2 is favourable. Superior values of VIP2 are observed for higher values of plasma parameters. The values climb with increasing V_{gs} and hit their peak values, after which there is a drop in the magnitude. By virtue of a more prominent fall in the values of g_{m3} compared to g_m, VIP3 has its paramount value found for highest value of plasma parameters i.e., longest channel length at higher values of V_{gs}. The peak value shifts towards higher V_{gs} upon increasing plasma parameters.

IMD3 also known as the third-order intermodulation distortion, represents the extrapolated intermodulation distortion power at which the first and third-order intermodulation power is equal [27]. IIP3 is defined as the third-order intercept input power characteristics. Higher values of IIP3 whereas lower values of IMD3 are preferred for improved linear characteristics. Figures 14 and 15 illustrate the impact of plasma parameters on IIP3 and IMD3 of the aforesaid device. IMD3 increases with increasing V_{gs} before attaining saturation for all investigated cases. Further, it is witnessed that lowest value of IMD3 is encountered at highest value of plasma parameters.

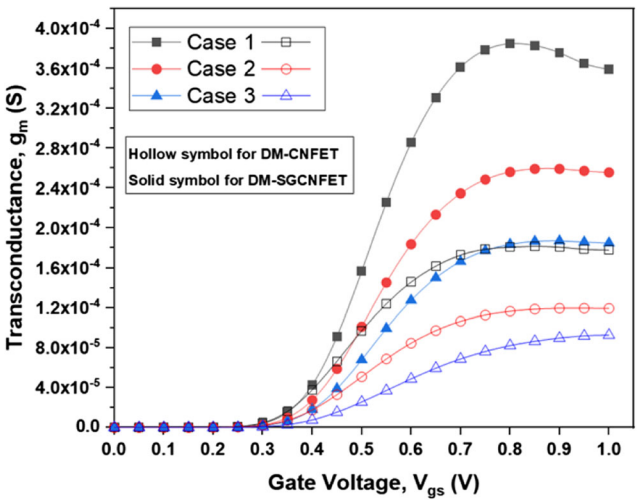


Fig. 9 Impact of plasma parameters variation (Table 3) on transconductance (g_m) at V_{ds} = 0.8 V

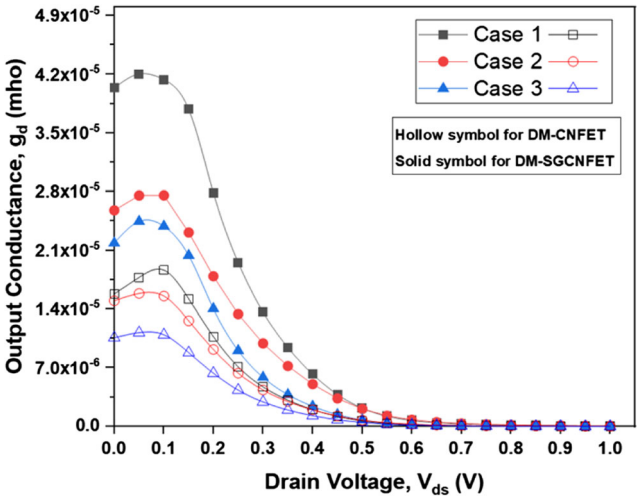


Fig. 10 Impact of different values of plasma parameters (Table 3) on output conductance (g_d) at V_{gs} = 0.5 V

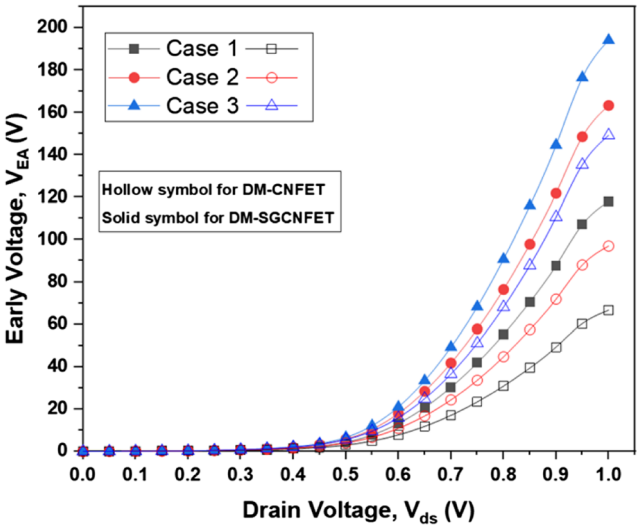


Fig. 11 Impact of different values of plasma parameters (Table 3) on early voltage (V_{EA}) at V_{gs} = 0.5 V

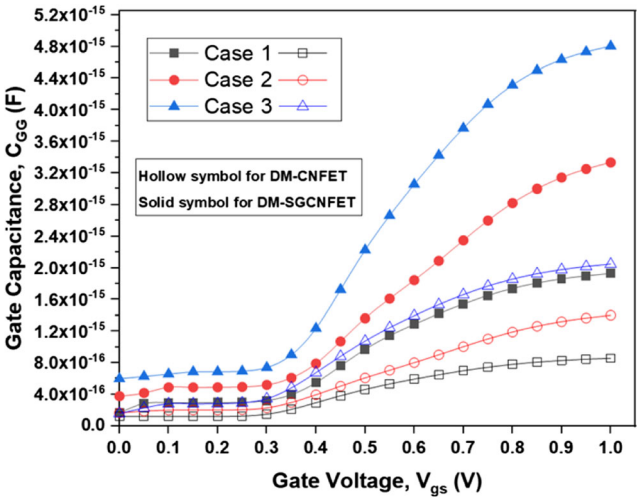
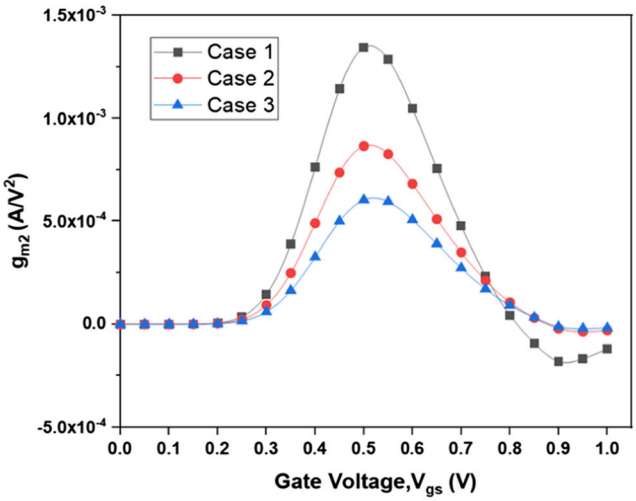


Fig. 12 Impact of different sets of plasma parameters (Table 3) on Gate Capacitance (C_{GG}) at V_{ds} = 0.8 V



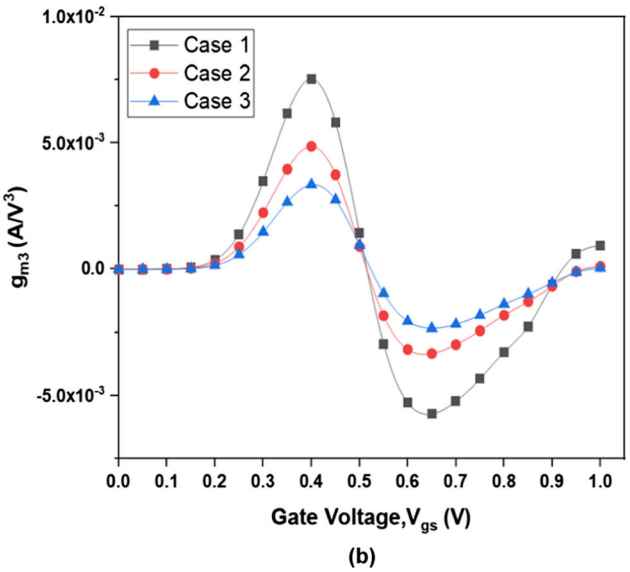
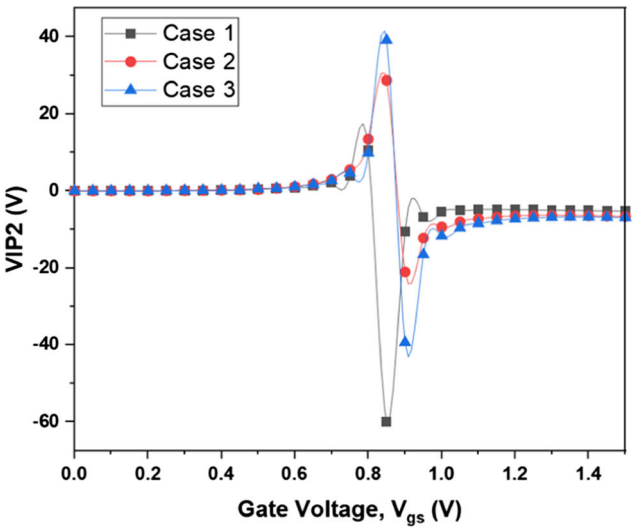


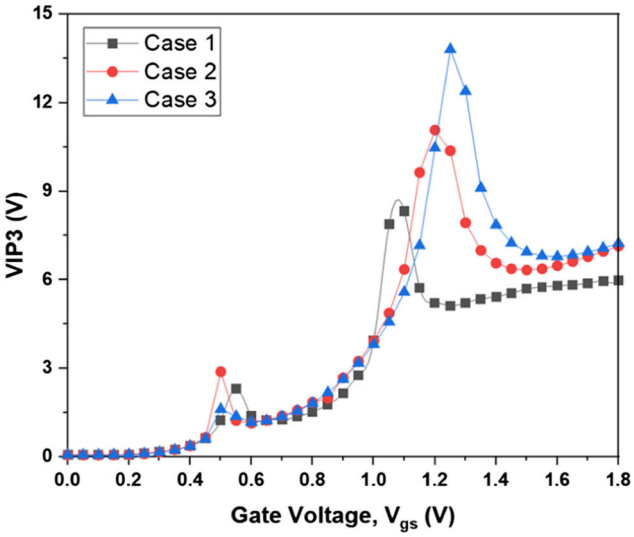
Fig. 13 Impact of plasma parameters variation (Table 3) on (**a**) g_{m2} (**b**) g_{m3} (**c**) VIP2 (**d**) VIP3 at V_{ds} = 0.8 V

Magnitude of the peak value of IIP3 increases with increasing sets of plasma parameters, i.e., for longer channel lengths, its peak value is higher. The peak values shift towards higher values of V_{gs} with increment in plasma parameter values. 1-dB compression point is defined as the input power at which output power gets deviated from linearity by 1 dB. The peak value is marked up for increasing sets of plasma parameters, ensuring high linearity which is represented in Fig. 16. Table 4 summarizes the variation in linearity metrics with increasing values of plasma parameters.

The obtained results further cement the DM-SGCNFET as a superior device for real life applicability. The results reported in this investigation can deeply benefit plasma-based nanoarchitectonics and related applications. DM-SGCNFET takes over the advantages of DM-CNFET and improves them even further through a high-k stacked oxide structure, thus



(c)



(d)

Fig. 13 continued.

stamping itself as a better device. Other applications of the simulated device include bio & electro-chemical sensors for detecting biological species, on-chip interconnect applications etc. [28–32].

4 Conclusion

The proposed device with SiO_2-HfO_2 Stacked Gate-Oxide structure has been proposed to counter the downscaling problems associated with conventional SiO_2 gate oxide. Researchers are propelling in the direction of novel materials for FETs instead of last mile optimisations of the current ones. Usage of high-k dielectrics fall in the same line. The simulated

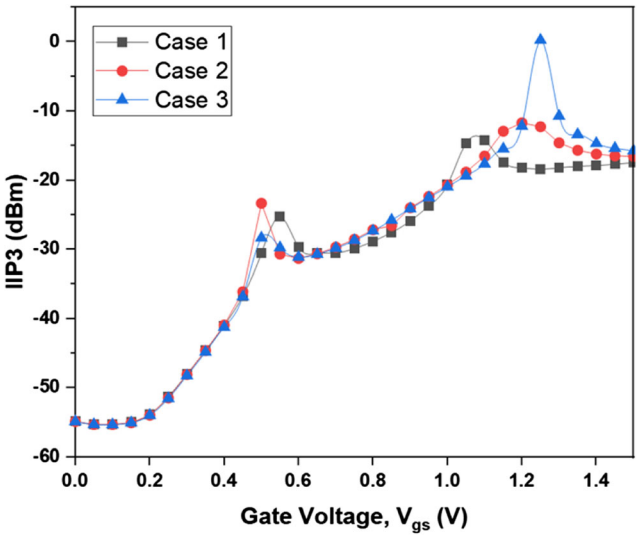


Fig. 14 Impact of plasma parameters variation (Table 3) on IIP3 at V_{ds} = 0.8 V

DM-SGCNFET device showcases much better electrical properties such as higher values of drain current, I_{ON}/I_{OFF} ratio, g_m, g_d, V_{EA}, gain and C_{GG} along with lower values of R_{ch} when compared against a SiO_2 based DM-CNFET device. Further, the linearity of the said device has been examined by comparing the linearity metrics of a transistor for varying sets of plasma parameters. From the simulations carried out in this study, it can be inferred that the preferred attributes (i.e., superior values of VIP2, VIP3, IIP3, 1-dB compression point and lower values of g_{m2}, g_{m3} and IMD3) are observed for higher values of plasma parameters.

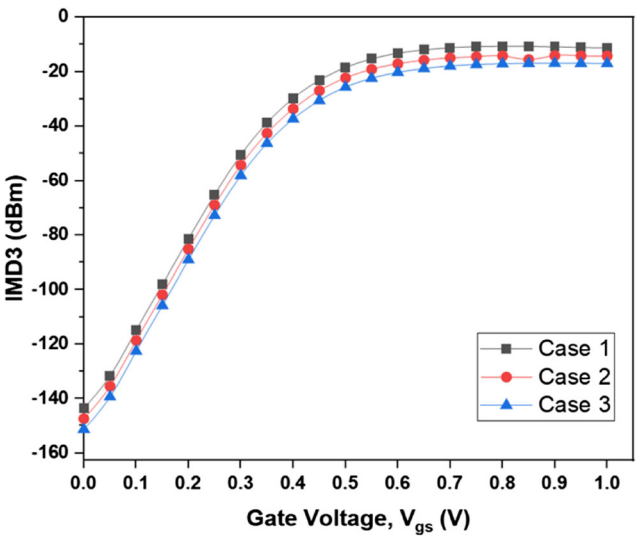


Fig. 15 Impact of plasma parameters variation (Table 3) on IMD3 at V_{ds} = 0.8 V

Table 4 Review Table

S.No	Linearity metrics	Trend with increasing values of plasma parameters (n_{oe}, T_e & n_{oi}, T_i)
1.	g_{m2}	Lower peak values
2.	g_{m3}	Lower peak values
3.	VIP2	Higher peak values
4.	VIP3	Higher peak values
5.	IIP3	Higher peak values
6.	IMD3	Lower peak values
7.	1-dB Comp. point	Higher peak values

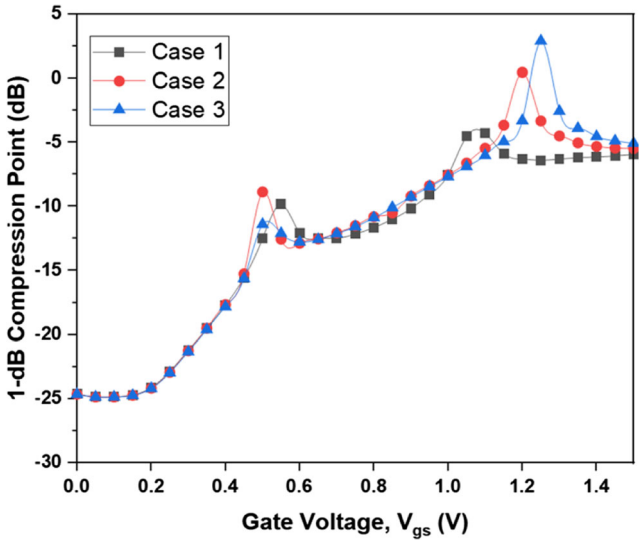


Fig. 16 Impact of plasma parameters variation (Table 3) on 1-dB Compression Point at V_{ds} = 0.8 V

Acknowledgements Author Mansha Kansal is thankful to Plasma and Nano-Simulation Lab, Delhi Technological University, for assisting this research work.

Availability of Data and Material The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Authors' Contributions Not Applicable.

Funding No funds, grants, or other support was received.

Declarations

Ethics Approval Not Needed.

Consent to Participate Not Applicable.

Consent for Publication Not Applicable.

Conflict of Interest The authors declare that they have no conflict of interest.

References

1. Gupta AK, Raman A (2020) Performance analysis of electrostatic plasma-based dopingless nanotube TFET. Appl Phys A Mater Sci Process 126(7):573. https://doi.org/10.1007/s00339-020-03736-7
2. Phokharatkul D, Ohno Y, Nakano H, Kishimoto S, Mizutani T (2008) High-density horizontally aligned growth of carbon nanotubes with co nanoparticles deposited by arc-discharge plasma method. Appl Phys Lett 93(5):053112 (1-3)
3. Ono Y, Kishimoto S, Ohno Y, Mizutani T (2010) Thin film transistors using PECVD-grown carbon nanotubes. Nanotechnology 21(20):205202
4. Qiu C, Zhang Z, Xiao M, Yang Y, Zhong D, Peng LM (2017) Scaling carbon nanotube complementary transistors to 5-nm gate lengths. Science 355(6322):271–276. https://doi.org/10.1126/science.aaj1628
5. Fahad HM, Hussain MM (2013) High-performance silicon nanotube tunneling FET for ultralow-power logic applications. IEEE Trans Electron Dev 60(3):1034–1039. https://doi.org/10.1109/TED.2013.2243151
6. Santolia I, Tewari A, Sharma SC, Sharma R (2014) Effect of doping on growth and field emission properties of spherical carbon nanotube tip placed over cylindrical surface. Phys Plasmas 21(6):063508 (1-8)
7. Li Y, Mann D, Rolandi M, Kim W, Ural A, Hung S, Javey A, Cao J, Wang D, Yenilmez E, Wang Q, Gibbons JF, Nishi Y, Dai H (2004) Preferential growth of semiconducting single-walled carbon nanotubes by a plasma enhanced CVD method. Nano Lett 4(2):317–321
8. Tewari A, Sharma SC (2015) Theoretical modeling of temperature dependent catalyst-assisted growth of conical carbon nanotube tip by plasma enhanced chemical vapor deposition process. Phys Plasmas 22(2):023505(1-8)
9. Kato T, Hatakeyama R (2010) Growth of single-walled carbon nanotubes by plasma CVD. J Nanotechnol 256906:1–11
10. Sharma SC, Gupta N (2015) Theoretical modeling of the plasma-assisted catalytic growth and field emission properties of graphene sheet. Phys Plasmas 22(12):123517(1-11). https://doi.org/10.1063/1.4938506
11. Gupta R, Sharma SC, Sharma R (2017) Mechanisms of plasma-assisted catalyzed growth of carbon nanofibres: a theoretical modelling. Plasma Sources Sci Technol 26(2):024006
12. Choi YC, Bae DJ, Lee YH, Lee BS, Park GS, Choi WB, Kim JM (2000) Growth of carbon nanotubes by microwave plasma-enhanced chemical vapor deposition at low temperature. J Vac Sci Technol A 18(4):1864–1868
13. Pang CS, Hwu JG (2014) Improvement in the breakdown endurance of high-κ dielectric by utilizing stacking technology and adding sufficient interfacial layer. *Nanoscale Res Lett* 9(1):464.i. https://doi.org/10.1186/1556-276X-9-464
14. Rewari S, Nath V, Haldar S, Deswal SS, Gupta RS (2017) Hafnium oxide based cylindrical junctionless double surrounding gate (CJLDSG) MOSFET for high speed, high frequency digital and analog applications. Microsyst Technol 25(5):1527–1536. https://doi.org/10.1007/s00542-017-3436-3
15. Manoj CR, Rao VR (2007) Impact of high-k gate dielectrics on the device and circuit performance of nanoscale FinFETs. IEEE Electron Device Lett 28(4):295–297. https://doi.org/10.1109/LED.2007.892365
16. Srivastava SK, Shukla AK, Vankar VD, Kumar V (2005) Growth, structure and field emission characteristics of petal like carbon nano-structured thin films. Thin Solid Films 492(1–2):124–130. https://doi.org/10.1016/j.tsf.2005.07.283
17. ATLAS User's Guide, SILVACO Int., Santa Clara, CA, USA, Version 5.26.1.R; (2018)

18. Kansal M, Sharma SC (2022) Plasma-based nanoarchitectonics for vertically aligned dual-metal carbon nanotube field-effect transistor (VA-DMCNFET) device: effect of plasma parameters on transistor properties. Appl Phys A Mater Sci Process 128(1):1–11
19. Yang X, Mohanram K (2011) Modeling and performance investigation of the double-gate carbon nanotube transistor. IEEE Electron Device Lett 32(3):231–233. https://doi.org/10.1109/LED.2010.2095826
20. Sathishkumar M, Samuel TSA, Vimala P, Nirmal D (2021) Performance analysis of HfO_2-SiO_2 stacked oxide quadruple gate tunnel field effect transistor for improved ON current. Silicon. https://doi.org/10.1007/s12633-021-01394-9
21. Sharma SC, Tewari A (2011) Effect of plasma parameters on growth and field emission of electrons from cylindrical metallic carbon nanotube surfaces. Phys Plasmas 18(8):083503
22. Ostrikov K, Xu S (2007) Plasma-aided nanofabrication: from plasma sources to Nanoassembly. Wiley-VCH, Germany. https://doi.org/10.1002/9783527611553
23. Mohsenifar S, Shahrokh Abadi MH (2015) Gate stack high-κ materials for Si-based MOSFETs past, present, and futures. Microelectron Solid State Electron 4(1):12–24. https://doi.org/10.5923/j.msse.20150401.03
24. Rewari S, Nath V, Haldar S, Deswal SS, Gupta RS (2016) Improved analog and AC performance with increased noise immunity using nanotube junctionless field effect transistor (NJLFET). Appl Phys A Mater Sci Process 122(12). https://doi.org/10.1007/s00339-016-0583-9
25. Sze SM, Ng KK (2006) Physics of semiconductor devices. Wiley, New York
26. Toan HLM, Goswami R (2022) Impact of gate dielectric on overall electrical performance of quadruple gate FinFET. *Appl Phys A Mater Sci Process* 128:103. https://doi.org/10.1007/s00339-021-05210-4
27. Kamal M, Yadav DS (2021) Effects of linearity and reliability analysis for HGO-DW-SCTFET with temperature variation for high frequency application. Silicon
28. Graham AP, Duesberg G, Hoenlein W et al (2005) How do carbon nanotubes fit into the semiconductor roadmap? Appl Phys A Mater Sci Process 80:1141–1151
29. Franklin AD, Koswatta SO, Farmer DB, Smith JT, Gignac L, Breslin CM, Tersoff J et al (2013) Carbon nanotube complementary wrap-gate transistors. Nano Lett 13(6):2490–2495. https://doi.org/10.1021/nl400544q
30. Dang T, Anghel L, Leveugle R (2006) CNTFET basics and simulation. *International conference on design and test of integrated Systems in Nanoscale Technology, (*28-33). https://doi.org/10.1109/DTIS.2006.1708731
31. Brady GJ, Way AJ, Safron NS, Evensen HT, Gopalan P, Arnold MS (2016) Quasi-ballistic carbon nanotube array transistors with current density exceeding Si and GaAs. Sci Adv 2(9):e1601240. https://doi.org/10.1126/sciadv.1601240
32. Jing G, Hasan S, Javey A, Bosman G, Lundstrom M (2005) Assessment of high-frequency performance potential of carbon nanotube transistors. IEEE Trans Nanotechnol 4(6):715–721. https://doi.org/10.1109/TNANO.2005.858601

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

iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 9 **Issue:** XI **Month of publication:** November 2021

DOI: https://doi.org/10.22214/ijraset.2021.38800

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.429
Volume 9 Issue XI Nov 2021- Available at www.ijraset.com

Potential and Achievement of Renewable Energy sources in India

Ayush Bansal[1], Ayush Jha[2], J. P. Kesari[3]
[1, 2]*Department of Applied Physics, Delhi Technological University*
[3]*Department of Mechanical Engineering, Delhi Technological University*

Abstract: ***Pure and environmentally friendly energy is of significant interest these days. It is one of the key enablements in achieving the Sustainable Development Goals (SDGs), accelerating social progress, and raising the standard of living. This work seeks broad insights into the Indian renewable energy framework, policy, approach, and socio-economic challenges. These include grid-independent and specific activities such as 'Utility-Scale Power Generation, Potential Multiple Geographies in India' and the Government of India's current achievements in advancing renewable energy production. Although much progress and progress has been made in the last decade of alternative energy maturation and expansion, improvements can be sought to increase the use of solar energy for conventional energy sources in India.***
Keywords: Solar Panels, Solar Energy, Renewable Energy, Sustainable Development Goals, power generation, cleaner fuels

I. INTRODUCTION

India is home to 1.3 billion people and is projected to surpass China's population by 2025 and become the most populous globally. Is meeting the growing energy demand of an increasing population. India's total primary energy consumption doubled from 326.77 million tonnes of oil equivalent (MTOE) in 2002 to 753.3 MTOE in 2017. [4] Over the past decade, India has increased its renewable energy capacity, especially from solar and wind. There is a significant increase in this regard, representing a significant portion of the renewable product in the country. [4]

India is very close to an "energy revolution" where critical sectors of the Economy are increasing their technology to reduce carbon emissions while driving sustainable growth. [8] India has big potential for solar energy, which is undoubtedly the largest renewable energy supply among Asian countries. As the third-largest energy consumer globally, India wants a healthy mix of all industrial energy sources. [8]

The Indian Renewable power generation Sector is the 4th most attractive renewable energy market in the world. By 2020, India will rank fourth in wind power, fifth in solar energy, and fourth in renewable energy efficiency. -3% barren land, bioenergy - 3%, small hydro - 2.2% [2]

The installed renewable energy generation capacity has been booming over the years, with the CAGR being 17.33 percent between FY16-20. With increasing government support and a better economy, the world is more attracted to investors' perspectives. Renewable energy will play an essential role as India, which is expected to achieve 15,820 TWh by 2040, begins to meet its energy demand on its own. According to the Paris Agreement, the Government of India is reaching its one hundred and seventy-five GW target by 2022 to reach 227 GW of renewable energy capacity (including 114 GW of solar capacity & 67 GW of wind generation capacity). The government plans to set a renewable energy capacity of 523 GW (including hydro to seventy GW) by 2030. [7]

Energy demand in India is growing rapidly, and by 2030, India's total energy demand will more than double, but electricity demand will almost triple compared to today. In addition, current specific resources have a climate and limited potential. Therefore, another kind of pure and unlimited generation is inevitable.

As of July 2019, the share of RE in India's total installed capacity for power generation is 22.5%. India, which has abundant natural resources, has a vast potential to generate electricity through RE sources. Current technologies have made it possible to use these renewable resources more efficiently to generate electricity. Fortunately, India has abundant natural resources for the industrial production of electricity through renewable energy. Commercially available mixed renewable sources of electricity in India are wind, solar, micro-water, biomass, tidal, geothermal energy. [2]

India has tremendous energy potential, and despite the difficulties it faces, it is addressing the growing demand for power generation standards and meeting those aspirations through renewable energy sources. Electricity demand in India is growing at one of the fastest rates in the world due to growth and economic growth.

© IJRASET: All Rights are Reserved

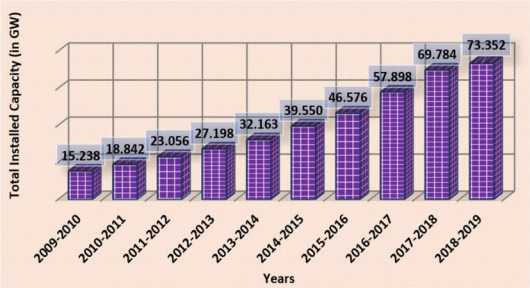


FIGURE 1. Growth of renewable energy over the past decade in India

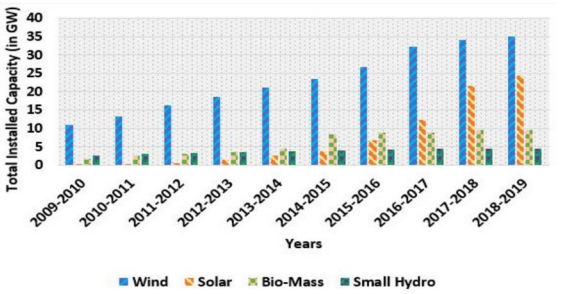


FIGURE 2. RE installed capacity (in GW) of various sources in recent years

To promote tremendous economic growth, it is necessary to switch to renewable cleaners, more cost-effective, and sustainable forms of energy from solar, wind, and nuclear sources. In fact, in recent years, we have seen a more drastic shift in government priority from promoting fossil fuel-based energy to solar and wind power through our various policies, programs, and incentives. Thus renewable energy is also at the forefront of empowering people at the grassroots level. The technology to implement alternative energy systems is widely available in India and needs to be brought in to meet the country's needs. [1]

II. RELATED WORKS

There have been many studies on India's opportunities and achievements in the field of renewable energy.

The study [2] covers India's perspective and perspective on solar energy consumption and undermines the country's potential for solar power generation. It covers schemes used by the government and essential measures to produce solar energy better. They also capture progress and incentives for solar energy production and consumption. This paper analyzes government policies, programs, and incentives for developing and developing the solar-wind sector in India and outlines potential limitations and solutions. It also reviews the development of an "offshore wind" that will be added to the nation's energy mix over the next decade. [4]

Another study is based on a comprehensive review of the policies of leading Indian states with renewable energy development, challenges, and international perspective. The Indian contrast in the global energy scenario is discussed in detail. In addition, barriers to the development of renewable energy production and policies of the Government of India to promote renewable energy production across India as well as globally are discussed in detail, as the challenges are similar to those of other countries.

Many news articles focused on India's recent achievements in the renewable energy sector in the past few years, which projected a healthy growth in renewable energy generation in the coming years/

III. DEVELOPMENT, ACHIEVEMENTS, AND INVESTMENTS

The Department for Promotion of Industry and Internal Trade (DPIIT) suggested data that showed that FDI inflows into the Indian non-conventional energy sector stood at US $ 9.83 billion between 2000 and 2020. More than US $ 42 billion has been invested. Has been investing in India's renewable energy sector since 2014. New investment in clean energy in the country reached US $ 11.1 billion in 2018. According to British Business Energy, an analytics firm, India ranks third globally in renewable energy investments and plans by 2020. Some of the significant investments and developments in the Indian renewable energy sector are as follows:

National Thermal Power Corporation Renewable Energy Limited (NTPC REL), a wholly-owned subsidiary of NTPC, has launched a tender for domestic manufacturers to build India's first green hydrogen filling station in Ladakh.
Reliance Industries has Rs. 750.00 crore ($ 10,070 million) of green energy.
Tata Power Solar Contract Rs. Rs 686 crore NTPC for the construction of 210 MW projects in Gujarat.
Adani Green Energy Limited (AGEL) has signed share purchase agreements with Softbank Group (SBG) and Bharti Group, planning to acquire a hundred percent stake in SB Energy India. India has a renewable energy portfolio of 4,954 MW, spread across four states.
The Central Electricity Authority has approved the expansion of the JSW Energy Karcham Wangtu hydropower plant from 1,000 MW to 1,091 MW (MW).

© IJRASET: All Rights are Reserved

GE Power India has completed the acquisition of 50% of the shares in NTPC GE Power Services Pvt Ltd. Ltd. agreed. For Rs.7.2 crore.
NTPC plans to commission India's largest floating solar power plant in Ramagundam in Telangana from May to June 2022. The total installed capacity is 447 MW.
The United States Agency for International Development (USAID) and the United States Finance Corporation for International Development (DFC) reported a $ 41 million loan guarantee program to help Indian SMEs invest in renewable energy.
Adani Green Energy announced plans to purchase a 250 MW solar power project in the northern state of Rajasthan (on behalf of Hero Future Energy). The estimated value of the deal is $ 10 billion ($ 136.20 million).
Adani Green Energy Limited (AGEL) has signed an agreement to acquire a 100 percent stake in Skypower Global's 50 MW solar power project. This increases the operational renewables capacity to 3,395 MW with a total renewables portfolio of 14,865 MW.
JICA (Japan International Cooperation Agency) signed a JPY 10 billion (US $ 90.31 million) loan agreement with Tata Cleantech Capital Limited (TCCL) to provide loans to companies in India for renewable energy generation and mobility solutions. Can bid. Moreover, energy saving to mitigate the effects of climate change by reducing greenhouse gas (GHG) emissions (in line with the principles of green credit).
India and the US agreed to restructure their strategic energy partnership to focus on clean energy sectors, including biofuels and hydrogen production.
India added 2,320 MW of solar capacity during the COVID19 epidemic.
The Solar Power Corporation of India (SECI) has started a massive central auction for solar parks and placed orders at 47 parks with a capacity of more than 25 GW.
The Adani Group is working to become the world's largest solar energy company by 2025 and the largest renewable energy company by 2030.

Chennai Railway Station is the first Indian railway station to reach the target of 100 percent daily energy demand through solar panels installed at the station platform shelters. SCR is first of the many Indian railway zones to adopt the idea of 'Energy Neutral' railway stations. To become the largest green railway network globally, Indian Railways is currently moving towards "net-zero carbon emissions" before 2030. To harness the gigantic potential of renewable energy, railways are increasingly moving towards solar-powered installations. Stations that contribute to environmental protection. [8]

Indian Railways already has a RE power capacity of 220 MW (MW) and a capacity of about 3,450 MW in the pipeline. The ministry says it has a capacity of twenty gigawatts of solar energy and aims to use its vacant land parcels to line up land-based solar plants for its traction power needs. A memorandum of Understanding(MOU) has been signed with the United Kingdom Foreign, Commonwealth, and Development Office (FCDO) for cooperation on energy and sustainability under its Power Sector Reforms Program. Indian Railways is also planning to integrate a type of solar plus storage hybrid pilot program with the Railway Traction Network in Dahod, Gujarat.

A report by the Ministry of Railways showed that one MW of solar capacity has been installed on the roofs of different stations and buildings by January 2021, namely: [8]

1) Three MW Roof Top Solar Plant at Howrah Railway Stations [8]
2) 1 MW Rooftop Solar at Katra Railway Stations [8]
3) 3 MW Solar Land-Based Project at Modern Coach Factory in Rae Bareli [8]
4) A 1.7 MW solar pilot project, was launched in June 2020 at Bina in Madhya Pradesh to power the Indian Railways' traction network directly through alternative energy. [8]

© IJRASET: All Rights are Reserved

IV. POTENTIAL

Over the years, the renewable energy sector in India has become a significant player in power generation capacity. Energy is one of the important components of economic growth and the well-being of nations. The existence and development of a suitable energy sector are essential for the Indian Economy's sustainable growth. India's energy sector is one of the most diversified in the world. The sources of energy production range from conventional sources like coal, lignite, natural gas, and oil to workable unconventional sources like wind, sun, water, and nuclear power. The country's demand for electricity has risen rapidly and is expected to continue to grow in the years to come; To meet the country's increasing demand for electricity, a massive expansion of the installed generation capacity is necessary. Renewable Energy Influence on the Indian Economy in the Last Five Years The renewable energy sector in India has seen tremendous changes in the political framework in recent years. The contribution to the total energy contribution in India. India ranks third out of 40 countries in the Renewable Energy Country Attractiveness Index 2016, with a strong government focus on promoting renewable energies. India has an approximate renewable energy potential of around 900 GW from sources such as wind (102 GW, bioenergy) 25 GW, small hydropower
20 GW, and solar power 750 GW. Renewable energies account for 15.90% of the total installed capacity in India. In March 2017, the installed capacity of renewable energies was 57,260 MW. Renewable energies have grown by more than 20% in the last five years. from 38,822 MW in 2015 to 57,260 MW in 2017. Wind power continues to dominate India's renewable energy industry, accounting for 29,151.29 MW in March 2017 compared to 25,088 MW in December 2015.
The installed capacity of renewable energies has increased steadily over time. Total installed capacity for renewable energies. The capacities in India grew steadily in the financial year 200715, from 9389 MW in 2007 to 34,351 MW in 2015 Wind and solar energy dominate the full potential of renewable energies in India, according to estimates of the states of the possibility of wind energy suggests suggesting that their potential is much higher in Gujarat and Tamil Nadu. In comparison, the prospect of solar energy indicates that its potential is much higher in Jammu and Kashmir, and Andhra Pradesh.

A. *Growth Drivers*

1) Through the Government of India, Ministry of New and Renewable Energy (MNRE), Generation Based Incentives (GBI), Capital and Interest Subsidies, Profitability Gap Financing (VGF). , Financing on favorable terms, tax incentives, etc.
2) The National Solar Mission is dedicated to promote the development and use of solar energy for power generation and other purposes with the ultimate goal of making solar energy competitive with fossil-based energy alternatives.
3) The National Solar Mission aims to reduce the cost of solar power generation through long-term action, large-scale implementation goals, aggressive RandD, and national production of raw materials, components, and critical products.
4) The government has provided a liberal environment for foreign investment in renewable energy projects. The establishment of its financial institution, the Indian Renewable Energy Development Agency (IREDA), provides new impetus to promote, develop and expand financial assistance for renewable energy. And energy-saving / efficiency projects.
5) Renewable energies are becoming very competitive than fossil energy generation, as the prices of solar modules have fallen by almost 80% since 2008.
6) The Reserve Bank of India (RBI) revised the guidelines for all scheduled commercial banks, including renewable energy in priority sector categories, in addition to the existing categories making significant progress for renewable energy in the priority sector of lending, as well as bank lending for rooftop solar as part of a Treat mortgage loan/home loan with a subsequent tax break.
7) Focus on the development of workforce qualifications: "SuryaMitra Scheme" was launched in May 2015 to create 0.05 million trained staff over a period of 5 years (201516 to 201920). In February 2017, almost 6653 SuryaMitra were trained under the program and created jobs. Opportunities for unemployed young people.

V. IMPACT ON HEALTHCARE SECTOR

The health care sector produces massive amounts of greenhouse gas emissions. The big machines and instrumentation that require to be operational all day long have enormous energy needs. Reducing the healthcare sector's emissions is a remarkable advancement in sustainability in healthcare. Renewable energy has the ability to facilitate the healthcare business lower expenses, leading to lower medical prices. Considering what amount of power the healthcare sector must operate daily, a marginal modification in energy bills would signify tangible savings. By reducing overhead costs with renewable energy sources, hospitals might be able to create very affordable patient care. Renewable energy sources can facilitate making health care more accessible. one of the significant essential problems. People in some parts of India still face is not having access to dependable energy sources.

© IJRASET: All Rights are Reserved

The result of that is usually inadequate healthcare provisions. These energy sources don't need a nexus to an electrical grid, creating energy more flexible. They'll conjointly help convey medical provides safely. For example, vaccines often require temperature-sensitive storage, which can be difficult to care for over long trips. Solar panels on transport equipment guarantee cooling and transportable solar cold storage units do not scatter electricity, keeping them safe as companies distribute them with zero power requirement. Renewable energy-dependent hospitals show improved performance. Analysis administered by the Council on Energy Environment and Water (CEEW), Supported by Oxfam India, evaluated the role of electricity access on health care consequences in rural India. The research suggested that, on average, health facilities with solar energy panels attended up to 50% additional out-patients every month, addressed up to 50% more institutional deliveries, accepted a more comprehensive range of in-patients, and enforced day-and-night services. This was correct even for health centers with electricity provided for fewer than twenty hours per day. In addition, staff from the health centers had positive perceptions of Such systems. For example, approximately 98% of employees reported lower disruptions in daily functioning, and 80% reported savings in electricity prices.

While shifting to renewable energy sources will help save the health care industry money, it's conjointly proven to help build the healthcare sector a lot of sustainability and lower its carbon footprint within the long run. In addition, renewable energy sources have proven to help cut back the rising cost of healthcare and empower more individuals to access the healthcare they need.

VI. GOVERNMENT INITIATIVES AND POLICIES

Indian authorities, through the Ministry of New and Renewable Energy (MNRE), are actively working to encourage the adoption of renewable energy sources by offering a variety of incentives such as product-based incentives (GBIs), capital and interest rebates, and possibilities. Gap funding (VGF), discount finance, monetary incentives, etc.

1) The National Solar Mission works with the final goal of creating sun power compete with fossil-based electricity options. Another intention is to lessen the cost of solar energy generation withinside the country via long-time period policy, large-scale deployment goals, competitive R&D, and the local manufacturing of critical raw materials, components, and products.
2) The status quo of a committed monetary institution – the Indian Renewable Energy Development Agency (IREDA), makes for renewed impetus at the promotion, improvement, and extension of monetary assistance for renewable electricity and power efficiency/conservation projects.
3) Reserve Bank of India revised the pointers for all scheduled industrial banks, which include renewable power in the categories priority area, further to existing categories making massive inroads for renewable power in the priority area lending, additionally financial institution loans for solar rooftop structures to be dealt with as part of home or home improvement loan with subsequent tax benefits.
4) Focus on ability improvement of the workforce: "SuryaMitra Scheme" was released in May 2015 to create 0.05 million educated employees within a length of five years (2015-sixteen to 2019-20).

A. *Sector Policies*

1) The mandate of SECI (Solar Energy Corporation of India) allows to carrying out of a wide range of functions to facilitate the implementation of the National Solar Mission. SECI aims to improve Renewable Energy (RE) technology and ensure inclusive RE power reform across India.
2) The Ministry of New & Renewable Energy (MNRE) released the National Offshore Wind Energy Policy, 2015, which authorize MNRE to discover and promote the deployment of offshore wind farms withinside the Exclusive Economic Zone (EEZ) of India, and the National Institute of Wind Energy (NIWE) has been authorized as the Nodal Agency for improvement of offshore wind power in the country and to perform allocation of offshore wind power blocks, coordination and allied capabilities with associated ministries and agencies. It is deliberate to install an introductory offshore wind energy venture off the Gujarat coast soon.
3) The Policy for Grid-connected Solar Roof-top Projects: Joint Electricity Regulatory Commission (JERC)/State Electricity Regulatory Commissions (SERC) of 29 States/UTs, has notified regulations/tariff order for grid-related sun rooftop projects.

State Programs: The State Electricity Regulatory Commissions in Andhra Pradesh, Haryana, Punjab, Madhya Pradesh, Maharashtra, Rajasthan, Tamil Nadu, Gujarat, Kerala, Punjab, Orissa, and West Bengal have introduced priority tariffs to purchase energy from wind power projects.

New Solar Policy in 2016 - Delhi, Himachal Pradesh, and Haryana.

New Solar Policy in 2015 - Telangana Jharkhand, Gujarat, and Andhra Pradesh

© IJRASET: All Rights are Reserved

VII. CHALLENGES

Renewable energies have grown significantly over the past decade, but there is still a long way to go to reach their full potential. Other obstacles/restrictions to the promotion of renewable energies are national and international guidelines, the availability of land for renewable power plants, political and social awareness as well as financial and technical considerations. As well as related guidelines and emphasis on economic energy consumption.
Some of the Major Barriers are:

A. *High Preliminary Cost*

1) The investment costs for installing renewable power plants are higher than for non-renewable power plants.
2) Prohibits consumers from adopting them as they prefer to keep primary costs low rather than lowering maintenance and operating costs.
3) The lack of access to favorable capital parameters such as inflation rate and interest rate during the installation of the renewable energy source also contributes significantly to the total acquisition costs.
4) The framework for the commitment to purchase renewable energy is not convincing, and most of the guidelines only apply to existing technologies.

B. *Economic and Financial Barriers*

1) Commercial banks are the ones that help finance RE projects. Although the banks offered their support, awareness did not spread, and the lack of information on many social power projects made it impossible for the banks to help.
2) Funding for renewable energy projects depends on issues such as project size, development, and technology.
3) If debt or equity demand is too high, Financing becomes expensive. Most banks are more willing to finance smaller projects than large projects.
4) Financing is provided through external Financing, syndicated loans, and bridging loans, among other things.

C. *Transmission and Distribution Losses*

1) In the transmission, distribution, and use of electricity, an average of about 30% of the losses was observed.
2) If the generation of renewable energies takes place far from the load center, the losses will increase sharply.
3) These losses should be monitored by the government, which is establishing a separate power distribution management department to allow for more efficient management and integration of future renewable energy generation centers into the grid.
4) As the network becomes more complex, so does the number of inverters, and this leads to high switching losses; This can be reduced by using bidirectional switches.

D. *Technical Barriers*

1) The main problem is that consumers are unaware of new technologies due to a lack of advertising. For example, solar technologies are misleading from the customer's point of view because they are available in abundance during the solar period, and they are not aware that some technologies make them very reliable in combination with hybrid photovoltaic systems.
2) Most of the generation of renewable energy is seasonal, and hydropower is dependent on rainfall and the availability of water.
3) The radiation that hits the globe depends on geographic locations, atmospheric conditions, are and the movement of the earth.
4) In some places, the data required to commission a system is not available, so there is a need in the country to set up stations that help to collect data on solar radiation and thus influence the progress of solar energy projects.
5) Communication between states on renewable energies should aim to overcome barriers. Hence, information support is essential to achieve the perfect articulation of efficient TERs, which will also propel and increase production.

E. *Research and development*

1) The research will require enormous government funding, but this is not currently available. In addition, companies that assume the market risk of introducing new technologies cannot take full advantage of their developments.
2) Technically trained candidates with specific management and development skills are unavailable due to insufficient training institutes.
3) There is insufficient guidance and technical support for engineers, resulting in inefficient use of renewable energy resources.

© IJRASET: All Rights are Reserved

F. Land Clearance Problems

1) Hydropower generation requires the evacuation of people from areas with enormous water potential for dam construction, leading to the destruction of the habitats and animals in that area.
2) Wind turbines are placed on a farm that represents the habitat of birds and also leads to deforestation. Most of the birds died due to the turbine blades, but some measures were taken to protect the birds by placing birds. Radars detect birds in this area and stop the turbine if the turbines pose a potential threat to the birds.
3) Building a solar park requires a considerable area, and the main problem is getting the radiation data and identifying the hotspots.
4) In biomass, there exists a risk with untested fuel supply and conversion technologies, and the production costs are very high compared to other production methods. Another problem is the lack of government support to give developers free land to build plants.

VIII. CONCLUSION

GOVERNMENT'S INITIATIVE	DATE
In July 2021, to encourage rooftop solar (RTS) throughout the country, notably in rural regions, the Ministry of New and Renewable Energy plans to undertake Rooftop Solar Programme Phase II, which aims to install RTS capacity of 4,000 MW in the residential sector by 2022 with a provision of subsidy.	July 2021
In July 2021, the Ministry of New and Renewable Energy (MNRE) gave the go ahead to NTPC Renewable Energy Ltd., a 100% subsidiary of NTPC, to build a 4,750 MW renewable energy park at the Rann of Kutch in Khavada, Gujarat. This will be India's largest solar park to be developed by the country's leading power producer.	July 2021
In June 2021, Indian Renewable Energy Development Agency Ltd. (IREDA) has invited bids from solar module manufacturers for setting up solar manufacturing units under the central government's Rs. 4,500 crore (US$ 616.76 million) Production Linked Incentive (PLI) scheme.	June 2021
As of March 2021, State Bank of India financed Rs. 319.18 billion (US$ 4.28 billion) in renewable energy projects in India, wherein the bank financed 752 renewable energy projects, with a total installed capacity of 13.8 GW.	March 2021
In June 2021, the Competition Commission of India (CCI) approved ReNew Power to exchange equity shareholding by its existing shareholders with shares of ReNew Global. Along with this, the CCI also approved a reverse triangular merger of ReNew Global's subsidiary with RMG II.	June 2021
In April 2021, the Central Electricity Authority (CEA) and CEEW's Centre for Energy Finance (CEEW-CEF) jointly launched the India Renewables Dashboard that provides detailed operational information on renewable energy (RE) projects in India.	April 2021
In April 2021, the Ministry of Power (MoP) released the draft National Electricity Policy (NEP) 2021 and has invited suggestions from all stakeholders such as Central Public Sector Undertakings, Solar Energy Corporation of India, power transmission companies, financial institutions like Reserve Bank of India, Indian Renewable Energy Development Agency, HDFC Bank, ICICI Bank, industrial, solar, and wind associations, and state governments.	April 2021
In March 2021, the Union Cabinet approved a Memorandum of Understanding (MoU) in the field of renewable energy cooperation between India and the French Republic.	March 2021
In March 2021, Haryana announced a scheme with a 40% subsidy for a 3 KW plant in homes, in accordance with the Ministry of New and Renewable Energy's guidelines, to encourage solar energy in the state. For solar systems of 4-10 KW, a 20% subsidy would be available for installation from specified companies.	March 2021
In March 2021, India introduced Gram Ujala, an ambitious programme to include the world's cheapest LED bulbs in rural areas for Rs. 10 (US$ 0.14), advancing its climate change policy and bolstering its self-reliance credentials.	March 2021
In the Union Budget 2021-22, Ministry for New and Renewable Energy was allocated Rs. 5,753 crore (US$ 788.45 million) and Rs. 300 crore (US$ 41.12 million) for the 'Green Energy Corridor' scheme.	March 2021
Under Union Budget 2021-22, the government has provided an additional capital infusion of Rs. 1,000 crore (US$ 137.04 million) to Solar Energy Corporation of India (SECI) and Rs. 1,500 crore (US$ 205.57 million) to Indian Renewable Energy Development Agency.	March 2021
To encourage domestic production, customs duty on solar inverters has been increased from 5% to 20%, and on solar lanterns from 5% to 15%.	March 2021
In November 2020, Ladakh got the largest solar power project set-up under the central government's 'Make In India' initiative at Leh Indian Air Force Station with a capacity of 1.5 MW.	November 2020
In November 2020, the government announced production-linked incentive (PLI) scheme worth Rs. 4,500 crore (US$ 610.23 million) for high-efficiency solar PV modules manufacturing over a five-year period.	November 2020
On November 17, Energy Efficiency Services Limited (EESL), a joint venture of PSUs under the Ministry of Power and the Department of New & Renewable Energy (DNRE), Goa, signed a memorandum of understanding to discuss roll-out of India's first Convergence Project in the state.	November 2020
In October 2020, the government announced a plan to set up an inter-ministerial committee under NITI Aayog to forefront research and study on energy modelling. This, along with a steering committee, will serve the India Energy Modelling Forum (IEMF), which was jointly launched by NITI Aayog and the United States Agency for International Development (USAID).	October 2020
India plans to add 30 GW of renewable energy capacity along a desert on its western border such as Gujarat and Rajasthan.	2020
Delhi Government decided to shut down thermal power plant in Rajghat and develop it into 5,000 KW solar park	2020
The Government of India has announced plans to implement a US$ 238 million National Mission on advanced ultra-supercritical technologies for cleaner coal utilisation.	2020
Indian Railways is taking increased efforts through sustained energy efficient measures and maximum use of clean fuel to cut down emission level by 33% by 2030.	2020

© IJRASET: All Rights are Reserved

REFERENCES

[1] Y. P. Chandra, A. Singh, V. Kannojiya, and J. P. Kesari, "Solar Energy a Path to India's Prosperity," J. Inst. Eng. Ser. C, vol. 100, no. 3, pp. 539–546, 2019, doi: 10.1007/s40032-018-0454-6.

[2] R. M. Elavarasan et al., "A Comprehensive Review on Renewable Energy Development, Challenges, and Policies of Leading Indian States With an International Perspective," in IEEE Access, vol. 8, pp. 74432-74457, 2020, doi: 10.1109/ACCESS.2020.2988011.

[3] Kiesecker, J.; Baruch-Mordo, S.; Heiner, M.; Negandhi, D.; Oakleaf, J.; Kennedy, C.; Chauhan, P. Renewable Energy and Land Use in India: A Vision to Facilitate Sustainable Development. Sustainability 2020, 12, 281. https://doi.org/10.3390/su12010281

[4] Singh, A.K., Idrisi, A.H. Evolution of Renewable Energy in India: Wind and Solar. J. Inst. Eng. India Ser. C 101, 415–427 (2020). https://doi.org/10.1007/s40032-019-00545-7

[5] Manna, U. (2017, June 27). Renewable Energy In India: Potential, Growth And Policies. Envecologic. Retrieved October 21, 2021, from https://envecologic.com/renewable-energy-in-india-potential-growth-and-policies/

[6] Kumar. J, C.R., Majid, M.A. Renewable energy for sustainable development in India: current status, future prospects, challenges, employment, and investment opportunities. Energ Sustain Soc 10, 2 (2020). https://doi.org/10.1186/s13705-019-0232-1

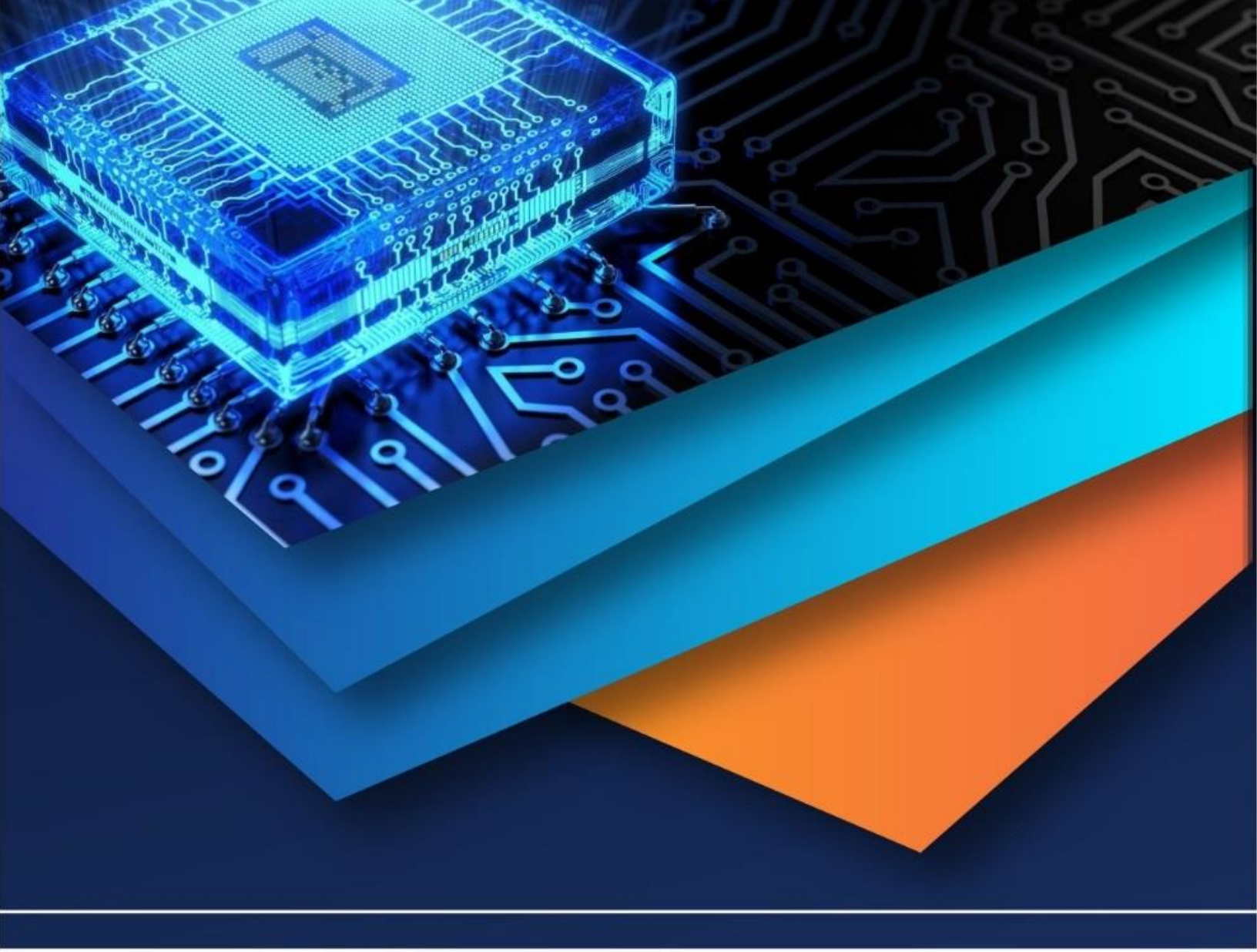
[7] Renewable Energy Industry in India: Overview, Market Size & Growth | IBEF. (2021, October 12). IBEF. Retrieved October 21, 2021, from https://www.ibef.org/industry/renewable-energy.aspx

[8] PBNS (2021, September 26). Green Railways: Chennai's MGR Railway Station gets powered by solar energy. NewsOnAIR. Retrieved October 21, 2021, from https://newsonair.com/2021/09/26/green-railways-chennais-mgr-railway-station-gets-powered-by-solar-energy/

[9] Ashutosh Verma. (2021, September 21). How Renewable Energy can transform healthcare. ETHealthworld.Com. Retrieved October 21, 2021, from https://health.economictimes.indiatimes.com/news/industry/how-renewable-energy-can-transform-healthcare/86380608

[10] Press Trust of India. (2021, October 12). India set to achieve 450 GW renewable energy installed capacity by 2030: MNRE. India Today. Retrieved October 21, 2021, from https://www.indiatoday.in/india/story/india-450-gw-renewable-energy-installed-capacity-2030-mnre-1863724-2021-10-12

© IJRASET: All Rights are Reserved



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089 (24*7 Support on Whatsapp)

Hindawi
Mathematical Problems in Engineering
Volume 2022, Article ID 1172050, 37 pages
https://doi.org/10.1155/2022/1172050

Research Article

Potential of a Nonperennial Tributary Integrated with Solar Energy for Rural Electrification: A Case Study of Ikukwa Village in Tanzania

Isaka J Mwakitalima,[1,2] **Mohammad Rizwan**,[1] **and Narendra Kumar**[1]

[1]*Department of Electrical Engineering, Delhi Technological University, Delhi, India*
[2]*Department of Electrical and Power Engineering, Mbeya University of Science and Technology, P.O. Box 131, Mbeya, Tanzania*

Correspondence should be addressed to Isaka J Mwakitalima; mwakitalimaisaka0@gmail.com

Received 20 August 2021; Revised 24 March 2022; Accepted 7 April 2022; Published 16 May 2022

Academic Editor: Alessandro Mauro

Copyright © 2022 Isaka J Mwakitalima 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.

This study evaluates the hydropower potential in the design of a micro-hydro/solar photovoltaic hybrid system with battery energy storage for increasing the access to electricity in Ikukwa Village in Mbeya Region of Tanzania. Usually, hybridized hydropower schemes are designed from perennial streams for the provision of electricity. This study incorporates the run-of-the river (COE) power scheme, which originates from the untapped potential of nonperennial hydro-energy source and the use of traditional approach of data measurements for Ikata tributary to design hybrid system. The system is optimized by the minimization of the total net present cost (NPC) and cost of energy (COE) using the soft computing method of Hybrid Optimization of Multiple Energy Resources (HOMER) software and artificial intelligent (AI) techniques. AI optimization techniques such as particle swarm optimization (PSO), grey wolf optimization (GWO), and GWO-PSO hybrid (GWO-PSOHD) algorithms have been employed for further optimal results. The data for solar radiation and the tributary have been obtained from the National Aeronautics and Space Administration (NASA) and traditional methods of measurements, respectively. The estimated maximum water flow rate and head are 2.943 m^3/s and 13 m, respectively. In the same period, the approximated theoretical power potential of the tributary is found to be 375 kW. Total NPCs obtained from HOMER, PSO, GWO, and GWO-PSOHD methods are $ 141, 397.76, $ 95 167.21, $ 92 472.82, and $ 91,854.10, respectively. Similarly, the optimal results of COE from HOMER, PSO, GWO, and GWO-PSOHD approaches are $ 0.1818/kWh, $ 0.1185/kWh, $ 0.1182/kWh, and $ 0.1181/kWh, respectively. Comparatively, PSO implementation has indicated the greatest energy cost, while the cost acquired by GWO-PSOHD is the lowest for all aforementioned AI optimization techniques. The tributary under study has a high potential of diversification of energy sources for rural electrification in the area of study and other parts of the world with comparable conditions.

1. Background

Energy is one of the greatest significant infrastructures needed for prosperous human life. Electricity improves the standard of living and economic growth of any country. Electricity supports the development of agriculture, industry, transport, tourism, and mining [1]. The majority of countries globally construct fossil fuel-driven power plants to meet energy demand through on- and off-grid energy systems. Sources of electricity are hydro, natural gas, and liquid fuels [2]. For the past at least twenty years, an estimated 80% of the Tanzanian power sector has been dominated by hydro-energy source [3, 4]. The total capacity of hydropower plants (HHPs) was equal to 561.84 MW. Capacities for off-grid NGPP and off-grid diesel power plants (DPPs) before were 25.25 MW and 28.682 MW, respectively. However, the addition of natural gas power plants (NGPPs) with the estimated capacity of around 495.44 MW surpassed the magnitude of hydropower source [3]. The approximated total capacity of electricity is 1450 MW of which NG, HPP, and NGPP correspond to 652.5 MW, 609 MW, and 188.5 MW, respectively [5]. Fossil fuels used in

electric power generation for grid extension are depletive and expensive. These fossil fuels are also detrimental to the environment causing pollution, global warming, and other numerous abnormalities [6]. The long gestation period of implementing the national electric grid and its power quality issues associated with the power delivery from distant generating places delays to avail electricity in off-grid locations [7]. Electrification in rural areas is lower than that in urban locations [3]. The decentralized approach via diesel-powered generators has been used to generate electricity in order to minimize the power deficit in the rural areas [5, 8]. However, diesel generators are also not preferable because of the high fuel price and carbon oxide emissions, which are harmful to the environment [8]. Therefore, it is advisable to include renewable energy sources in the power generation sector of Tanzania [5]. Renewable energy sources found in rural areas can be utilized to meet energy demand [9] since the country has a great potential of renewable energy resources, which can solve energy deficiency in the rural areas. These energy resources comprise hydro, wind, biomass, solar, geothermal, and tidal and wave [10, 11]. These renewable energy sources could have helped much to address power catastrophe in the nation, particularly in the sectors of agriculture, mining, tourism, industries, transportation, telecommunication, etc. However, the development of these nonconventional energy resources in the Tanzanian environment is hindered by several challenges such as shortage of human capital and training, large investment cost, limited awareness, minimum and limited research and development, centralized approach (monopoly) regarding the institutional framework, and more interest in centralized electricity from the national grid (TANESCO) leading to minimal adoption of decentralized power schemes [10].

Biomass, in form of firewood or charcoal, is one of the nonconventional energy resources, which is used extensively worldwide, especially in developing countries for food preparation and heating purposes. Biomass is an organic material derived from animals and plants that is a renewable energy source [12]. It is made up of sun-stored energy. The energy from biomass is transformed into heat when it is burned. The process of photosynthesis, which occurs when trees and plants develop, takes energy from the sun to transform CO_2 into carbohydrates such as starches, cellulose, and sugars [12] Carbohydrates are organic molecules found in biomass. Once trees (plants) die, the energy stored in carbohydrates is released by the breakdown process and emits CO_2 into the environment. Because the growth of new trees and plants replaces the supply, it is a renewable energy source. The use of renewable power is crucial because it promotes financial security, energy security, and ecological sustainability [12]. The transition from petroleum-based goods (i.e., fossil fuels) to substances derived from biomasses enhances the supply of long-term carbon-based feedstock. Some of the advantages of the biomass are as follows: (i) biomass energy is a renewable energy source that is derived from organic substances, plant, or animal waste, and these products are never depleted. Planting can be used to replace trees. (ii) It is healthier than fossil fuels; while burning, biomass emits CO_2, and it also collects CO_2 for its own development. However, CO_2 emissions from fossil fuels are discharged into the air and are damaging to the ecosystem. (iii) Biomass energy is a cheap and widely available form of energy. Biomass can also be a long-term sustainable power source if trees are replanted. (iv) It minimizes dumping heaps of garbage (landfills). By burning it for energy, biomass converts environmentally hazardous waste into something beneficial and (v) reduced the requirement of fossil fuels. The use of biomass as an alternative fuel source decreases our dependence on nonrenewable energy sources, which is better for the environment and more cost-effective [12]. Some of the drawbacks of using biomass are (i) high investment cost. A boiler for biomass has a higher initial cost than a standard gas or oil boiler, and (ii) it requires huge space. Its boiler systems are typically bigger than oil or gas burners and need a fuel storage area, (iii) damaging the environment. Because biomass uses human and animal wastes, methane gas emissions are hazardous to people and the environment, and (iv) it may result in deforestation and (v) provision of inefficient energy. The performance of biomass fuel is inferior to that of fossil fuel [12]. When compared to fossil fuels, biodiesel derived from biomass, i.e., ethanol, is relatively inefficient. It can lead to extinction, habitat loss, fragmentation, and other negative consequences for biodiversity [12]. The drawbacks of the biomass utilization can be further expounded as follows: nonrenewable energy sources are good standard goods that can be easily integrated into a wide range of mature power conversion systems. Biomass, on the other hand, has lower energy content, owing to its higher oxygen content. Furthermore, because biomass is a relatively light and dense substance, the quantity of biomasses to be handled might be significant, posing storage and transportation challenges [13]. The usage of fuels with high moisture content reduces a power plant's total energy efficiency due to the need for power for hidden heat of vaporization. In addition to diminishing the overall heating value, high moisture causes a rise in transportation and storage costs, affecting the returns and the usability of biomass-driven power stations, particularly at higher power scales [13]. Biomass drying, on the other hand, provides little advantage for improving qualities such as oxygen concentration, poor grinding capability, and hygroscopic behaviour (absorption of moisture content) [13]. As a result, pretreatment solutions must frequently be explored to improve the features of biomass as a fuel, making large-scale exploitation of these resources for bioenergy production uneconomic and inefficient [13].

The direct utilization of biomass-fired technologies presents several technological challenges, such as combustor fouling and corrosion induced by the alkalis of biomass wastes. Slagging and fouling impede heat transfer at combustor substrates and cause erosion and corrosion issues, reducing the equipment's lifetime [13].

Lignocellulosic biomass plays a great role in the biomass energy technology. Lignocellulosic biomass refers to the biomass originating from cellulose, which is cheap and easily accessible for producing energy [13, 14]. Forestry remnants, cellulosic wastes, agricultural remnants, and energy crops are all good sources of lignocellulosic biomass. The global

annual supply of lignocellulosic biomass is 181.5 billion tonnes [14]. The expense of this biomass is expected to range from US dollars 24 to US dollars 121 per tonne, type of crop, yield, area, and technique of analysis [13]. The total estimated energy yielded from the lignocellulosic biomass accounts for approximately 10% of world energy requirements [14]. Agricultural and forest leftovers alone contributed 30 EJ, a significant amount, to the yearly energy consumption of 4500 EJ [14]. There is biomass availability potential in Europe ranging from 615 to 728 million tonnes. In Italy, agricultural residues amount to approximately 12.8 million tonnes per year. 9.3 million tonnes can be generated from herbaceous crops (sunflower stalks, straw, and maize stalks) grown mostly in the north, and 3.5 million tonnes can be generated from arboreal crops (olive and fruit trees and vines) grown primarily in the south [13].

Tanzania possesses 33 million hectares of wood fields and forests, as per United Nations Food and Agriculture Organization's Global Forest Resources Assessment Study, but it losses approximately 400,000 hectares of forest each year [14]. Around 75% of harvested wood is not accounted for in government budget systems, resulting in revenue loss. The average household consumes approximately 47 kg of charcoal each month. Approximately 2500 million people worldwide rely on firewood for cooking and heating [10]. In the SSA region, including Tanzania, the population uses approximately 85 percent of all firewood that is taken from woodlands as energy or charcoal for cooking and heating [10]. For instance, as of 2018, the population was 56.32 million people. Only 35.6 percent of the overall population has electricity access; this corresponds to almost 36 million Tanzanians who do not have access to power. Biomass (e.g., firewood and charcoal) is still used as a source of energy by 80% of Tanzanians [10]. Whether burned inside or outside, the smoke created is a pollutant connected to negative health outcomes; particularly, respiratory disorders can drastically affect children and women [10]. Those who live in Tanzania's cooler southern highland regions frequently cook indoors, increasing their exposure to smoke pollution. Regardless of the high potential biomass in the country, there are only three power plants fuelled by biomass having at least installed capacity of 5.1 MW up to 18 MW, while the agro-industry-based power plants generate around 58 MW of their own electricity [10, 14].

Biofuels, principally bio-oil, bio-ethanol, bio-oil, and gasoline, and chemicals, can all be made from lignocellulosic resources [14]. For lignocellulosic biomass conversion, various conversion processes are available that take thermochemical, thermal, and biological routes [14]. The combustion of biomass for electricity and heat generation is referred to as thermal conversion. The process of converting biomass using extreme hotness and chemicals is termed as thermochemical transformation [14]. It is further classified into methods such as gasification and pyrolysis. These methods produce bio-oil and gasoline, as well as certain chemicals. Bioconversion is required for turning biomass into biofuels using microorganisms. Its primary product is bio-ethanol, although it also generates biobutanol, biogas (methane), and little chemicals [14].

Before bioconversion, pretreatment is essential to separate the ingredients of biomass for optimal fermentation [14]. Nonetheless, using feedstock necessitates a significant investment in power, heat, and work throughout the conversion process. The core of the problem is the costly pretreatment stage, which raises the entire cost of bioconversion. Chemical and physicochemical pretreatments are now the most effective. They are, however, harmful to the environment and produce hazardous by-products such as furfural. As a result, environmentally friendly biological pretreatment techniques are used on occasion, but with reduced fermentable sugar yields [14].

Integrated biorefineries increase the consumption and use of biomass to generate communal energy by mixing fundamental conversion, split-up, and downstream improvement processes. It is a necessity for fostering sustainable energy in a cost-effective manner. However, implementing them on a commercial basis is difficult. Recent research has focused on lignin chemistry and valorization, cost-cutting measures, sustainability of pretreatment procedures, advantages and disadvantages of various pretreatment methods, and transition processes of lignocellulosic biomass [14].

The conversion of lignocellulosic biomass into bioenergy is dependent on various processes, comprising (i) thermal transformation by burning, (ii) thermochemical transformation through both gasification and pyrolysis, and (iii) physiological (microbiological) transformation involving the use of biocatalyst [14].

Thermal Conversion: high-temperature direct combustion of lignocellulosic biomass produces thermal energy and ash particles. The thermal energy obtained can be used to generate electricity and power [14]. On a commercial basis, it is a low-cost method that has been thoroughly developed. However, it is not an environmentally favourable process because ash particles cause significant pollution [14]. Whole process is fully described as follows: direct combustion of biomass at hot temperatures (800–1600 C) produces thermal energy. When the fuel combines with oxygen, heat energy is produced, as well as flue gas containing CO_2 and water. The intensity of the heat flame can surpass 1650 C and is determined by the moisture contained in the biomass and heating value of the energy, the structure of the furnace, and the fuel-to-air ratio [14]. Because of their increased efficiency, biomass pellets are extensively used in commercial combustion in European countries. For this technique, biomass with moisture contained in biomass below 50% is recommended [14]. Combustion is carried out on a variety of scales utilizing a variety of equipment such as stoves, boiler, furnaces, turbo generators, and turbines. It is required for turning the chemical energy in biomass into mechanical energy. The combustion of fossil fuels results in massive emissions of CO_2, NOx, PM (particulate matter), and cinders. These discharges to the environment represent significant greenhouse gas risks [14].

Thermochemical Conversion: the conversion of biomass into liquid or gas, which is then upgraded into biofuel, physical catalyst, or heat, is used. This method necessitates a high level of energy consumption. The chemicals utilized are

frequently costly, and as a consequence, inhibitors are produced. The cost of cleaning up emissions and pollutants is inextricably linked to the total price of the process [14].

*Gasification i*s a thermochemical activity that combines the interaction of biomass in the presence of air, oxygen, or steam to generate a combination of gases, including hydrogen (H_2) methane (CH_4), carbon monoxide (CO), carbon dioxide (CO_2), nitrogen (N_2), and hydrocarbons called as syngas, producer gas, or synthetic gas, the quantities of which vary [14, 15]. The primary goal of the process of gasification is to maximize the yield of vapors while minimizing the production of unreacted char and hydrolysable hydrocarbons. There are basically three significant types of gasification processes: fixed bed, fluidized bed, and entertained flow [14]. Gasification generates heat that can be used by kilns or boilers. Because unattractive compounds such as tars are destroyed during gas combustion, the syngas used is barely cleansed. If the tar is not overwhelmed and the majority of the particulate matter is removed from the gasifier, syngas can also be used in combustion engines [14, 15]. Gas turbines encourage high-efficiency integrated gasification-combined cycle power, while good gas cleaning is required. Syngas is produced in a gasifier either explicitly or implicitly supplying temperatures of 600°C to 1000°C to the gasifier [14]. A successful gasification reaction has interval of time of around 3 to 4 seconds. Syngas is essential for the production of a wide range of chemicals and fuels such as organic acids, alcohol, esters, and hydrocarbon fuels [14]. Gasification has a broad range of applications being required for hydrogen production, thermally powered generation, and the synthesizing of chemicals and fuels [14]. This promotes energy industries based on gasification that generates an assortment of various chemicals and energy. Utilization of micro-turbine-driven biomass energy systems (syngas) and combined heat and power (CHP) gasification-based systems are examples of thermally powered generation [13].

Pyrolysis: pyrolysis is defined as a thermochemical process that can occur as an early step to combustion processes or as a standalone process for biofuel production. Pyrolysis is the degradation of lignocellulosic materials into an aromatic and carbon-rich solid in the lack of oxygen at temperatures starting from 300–900°C. This procedure yields biochar (solid type), bio-oil (liquid type), and flue gases (gaseous condition) [14].

Bioconversion is the process of converting biomass into biofuel by employing biocatalysts either directly or indirectly as part of a prior pretreatment step [14, 15]. Complex carbohydrates are broken down into intermediate sugars, which are then fermented to make biofuels such as ethanol. Biocontrol agents, such as enzymes or a cluster of microorganisms, enable this conversion. It is an extremely selective, high conversion efficiency method. Unfortunately, this procedure takes more time plus an extremely high pretreatment cost, indicating the process' infeasibility [14].

Direct microbial conversion (DMC) or consolidated bioprocessing (CBP) is a method of producing value-added products by combining enzymatic hydrolysis and fermentation using a cluster or a sole organism. Without any pretreatment, lignocellulosic biomass is transformed straight into biofuels [14].

Indirect Biological Conversion: lignocellulosic biomass is first decomposed into fermentable sugars, which are then fermented into biofuels [14].

Geothermal energy is the energy (in form of heat) that is produced and retained in the earth. It is both clean and sustainable, and it can be found in the bedrock and fluids underneath the earth's crust. It is available from low ground to several miles beneath the surface and even deeper to the exceptionally hot molten rock known as magma [12].

Some merits of geothermal energy are as follows: (i) it is ecologically friendly: a geothermal power station has a very low carbon footprint. The CO_2 emitted by geothermal power stations is one-eighth of that emitted by an equivalent coal power station. (ii) Stable resource: the plant's power output can be forecast and is not subjected to the equal low energy variations as solar or wind. (iii) A renewable resource: geothermal reservoirs are refilled naturally. (iv) No propellant involved: after construction, no mine or logistic activity is required. (v) Minimal land footprint: its plant installation requires a small amount of land. (vi) Economic criteria: some places are cost-competitive. (vii) Accessibility: most places have some amount of geothermal power available [12].

Some demerits of geothermal are as follows: (i) production of hazardous emissions: greenhouse gases (sulfur, CO_2, silica) are commonly found near plants. Toxic heavy metals such as mercury, arsenic, and boron may be present in the reservoirs. (ii) Geographical restrictions: suitable geothermal reserves are few. Prime locations are frequently located distant from major population centres. (iii) Pump powering costs: geothermal heat pumps require a power source. (iv) Expensive electricity: total costs for a megawatt geothermal power station typically range between $ 2 and $ 7 million. (v) Surface instability: the plant's construction may have an impact on the land's stability. (vi) Distribution costs: if energy is carried across great distances, the cost can become prohibitively expensive. (vii) Operation without a requirement of steam: if temperature is not properly managed, it can result in melting or other problems where energy is not dispersed or utilized effectively [12].

It has been expected that the global consumption of geothermal energy grows up to 200 TWh in the last two years (i.e., 2020) [16]. The annual growth rate of the consumed energy was estimated to be equivalent to 7%. The United States, Italy, Indonesia, Philippines, and New Zealand lead the world in geothermal power generation [10].Tanzania has significant geothermal energy resources, but they have yet to be completely measured [10]. The assessment of geothermal energy resources began in 1976 and Tanzania now has a potential of around 5000 MW capacity that can be generated by geothermal power resources, with the majority of prospects situated in the East Rift Valley System in Africa [10]. Tanzania's geothermal blessings and prospects are divided into three major zones: the northeastern zone, which includes the some areas of Kilimanjaro, Arusha, and Mara;

the southwestern zone, which particularly includes the regions of Mbeya and Rukwa; and the eastern parts of coastal belt zone, which includes the Rufiji Basin and is associated with magmatic and rifting encroachments [10].

The intensity of the geothermal power source can range from 60°C to 350°C. Power stations are not cost-effective for temperatures under 80°C owing to the reduced overall system efficiency [14]. Temperature is used to categorize geothermal resources for different applications such as conventional electricity generation, binary cycle electricity generation, and heating (heat pump). The possible use of geothermal energy can vary depending on the level of temperature of the geothermal fluid, geological pattern, and hydrogeological features [17, 18]. Different geothermal technologies are used either directly or indirectly for different applications such as electric systems, district heating systems in cold zones, and cooling systems in hot zones [17–20]. Direct applications of heat derived from geothermal energy resource are beneficial to environment and are now replacing traditional sources of energy (fossil fuels) [18]. Possible applications of geothermal energy resources depending on the fluid temperature are such as electricity production at temperature above 90°C; direct Applications of thermal energy for industrial process, agricultural production, and space heating at temperature below 90°C; and exceptional operations at fluid temperatures lying between 30 and 50°C [17, 19]. Geothermal fields possessing low temperatures have been harnessed for industrial applications in China, Hungary, Iceland, Hungary, Germany, Russia, Turkey, France, and other nations for a number of decades, and many authors have recently proposed the use of geothermal energy alone or in conjunction (integrated) with several other nonconventional energy sources [19]. For example, in Kamchatka, Russia, geothermal field with low-temperature field of meteoric origin has harvested thermal water since 1966, mostly in the manner of artesian flow, to supply the multiple swimming pools, the district heating with two settlements, greenhouse gardening, and fish breeding [19]. Enhanced Geothermal Systems (EGSs) are man-made geothermal tanks formed by hydraulic fracking and other stimulation techniques in ultralow permeability rocks. The principal purpose is to foster favourable circumstances for the circulation of water and heat mining in order to support surface energy uses such as district heating, power production, and straightforward industrial heat use [17, 20].Technological advancements in hydraulic fracturing and multistage hydraulic fracturing have revolutionized flow permeability shale in recent years. Even though the high price remains a significant obstacle, it is hoped that this technology will be used to design fracture structures with a vast surface area of EGS (20). Since the 1970s, many efforts have been made to construct cost-effective EGSs. For example, the EGS at Fenton Hill in New Mexico was the first effort to harvest heat from porous rock. Multiple vertical hydraulic fissures were built to connect and allow circulation of water between two wells (holes) apart, extracting heat from 190 degree Celsius at 3500 metres depth. Projects of a similar nature have been carried out in Korea, Australia, and Europe. With hurdles from seismicity, connection, and financial feasibility, these have met with uneven success. Another recent project is the thermal field known as Qiabuqia in China, which has been dug to a depth of 3.7 km depth [20].

There are numerous classifications for geothermal energy resources, based upon enthalpy or temperature (basic types are dry steam energy plants, flash steam energy plants, and binary ORC (organic Rankine cycle) or the concept of exergy) [14, 16]. It is primarily classified into three types based on the temperature or enthalpy, namely low, medium, and large enthalpy resources. The major vocabulary used for classification based on exergy is specific exergy rate (SER) or specified exergy index (SEI). Geothermal resources are divided into three categories based on SER or SEI: low, high, and intermediate exergy resources. Because it is not possible to determine the type of fluids based solely on temperature or enthalpy, SEI-based categorization of geothermal energy is also necessary [16]. For saturated steam and water, the SEI value ranges from 0 to 1 [14].

Geothermal resources can directly be utilized for heating purposes. Geothermal power can be used to generate power, desalinated water, and space cooling [17, 21]. Also, apart from the basic kinds of power generation via geothermal energy resource (i.e., dry steam energy plants, flash steam energy plants, and binary ORC) [14, 16, 21], it can also be integrated by other energy resources: *incorporation of geothermal power system in conjunction with other energy sources*: cascading refers to the pairing of various power-producing systems. Geothermal power nowadays is currently combined with solar, biomass, thermoelectric generators (TEGs), trigeneration, hydrogen production, and other kinds of direct use [21]. When geothermal resources are extracted, there is frequently residual enthalpy in the geo-fluid that might be utilized in a new application. A moderate temperature from geothermal deposit can produce power in an ORC power plant, where the heating rate of the geothermal saline is sufficiently high to be utilized in heating (heat pump) process [16].

Geothermal and Solar Photovoltaic (PV) Hybrid Systems: hybrid solar geothermal energy systems are characterized by their simplicity to set up in a variety of operating circumstances. They have been demonstrated to be reliable by employing low-cost collectors to set prices comparable with other kinds of electricity. There is also the possibility of multi-generation (heating/cooling) applications, as well as significant capital cost reductions through shared equipment within systems. The drawbacks of hybrid geothermal power systems include reduced efficiency cycles, rusting and scaling difficulties, and multi-generation facilities that necessitate numerous heat sinks [19, 21].

Other Geothermal Energy-Based Hybrid Systems: while the majority of research for geothermal-hybrid systems has focused on pairing geothermal and solar technologies, there has been significant research into systems that use geothermal technology in conjunction with organic material (biomass), hydrogen production, TEG, and cogeneration with additional freshwater production [16, 19].

Improved Geothermal Systems: drilling to a depth where the mantle's temperatures are adequately hot to transfer heat from the permeable rocks into a working fluid is required for enhanced geothermal systems (EGSs) [16, 20]. The fluid is then directly discharged, and energy is generated by any of the existing power plant types. Hot dry rock is relatively prevalent in the planet with average thermal gradients ranging from 25 to 30 °C/km deep. Some preliminary thinking suggests employing thick fluids to boost circulation pathways and lower rock permeability, allowing for smoother circulation starting from the injection well into the extraction well [16, 20].

Before drilling and designing geothermal plant, knowing scientific-technical data connected to the possibility of geothermal resources and the sustainability of utilization is a critical task for assessing the economic feasibility of geothermal exploitation and the creation of future project plants. More information such as the applicability of system periphery conditions during long-term utilization, the interference of fluid extraction and reinjection during production, and the efficacy of geothermal production systems can be gathered in this manner. Numerical simulation is recognized as a critical instrument for the development and evaluation of geothermal energy, which is otherwise regarded as highly dangerous [19].

As it has explained above, the development of renewable energy technology in Tanzania is at preliminary stage [5]. However, micro-hydropower and solar PV have been implemented in off-grid locations of the country [5]. Hydropower generation technology is mature, and due to its high potential of hydro-source, several projects of various scales are being implemented off-grid locations [11, 22, 23]. Geographically, Tanzania is located close to the equator, and therefore, it receives large amount of solar radiation in almost all parts of the country. The central region has highest potential of solar energy [24, 25]. The chief purpose of this work is to execute feasibility study and make validation by designing hybrid system using micro-hydropower and solar PV technologies. Therefore, the availability and high potential of hydro- and solar energy sources in Tanzania can be used to design decentralized electricity generating systems including optimal hybrid power plants for rural areas. Similar studies have been performed locally and internationally concerning optimization of hydro-solar PV hybrid electric generating systems for off-grid locations. Some of these studies have mainly used hydro-energy source from perennial rivers, water streams, and agricultural irrigation channels to conduct feasibility studies to assess the feasibility of generating electricity for off-grid locations [26–28]. Different research methods and techniques have been used for the simulation [29–31]. Kalla et al. proposed the hybrid micro-grid system by integrating micro-hydro-turbine system, wind, solar PV, and batteries using a 1~voltage source inverter. The target of the designed system was to balance the electrical energy source among the three sources using real-time approach termed as digital signal processor controller (DSPC). The results indicated that the grid voltage and frequency are maintained constant regardless of expected variations in loads and intermittency from the penetration of wind and solar PV [32]. Li et al. optimized micro-hydropower plants/solar PV/pumped hydro-storage hybrid system and devised the day-ahead best performance strategy of the pumped hydro-energy storage (PHES) to balance the generated power curve and the consumed electrical load while maximizing the income of the storage. Optimization through chance constrained programming (CCP) and multiple objective PSO (MOPSO) was also done, and further the day-ahead optimal performance strategy, the pumped hydro-energy storage (PHES) balanced the generated power curve and consumed electrical load while maximizing the financial income of the storage. The authors investigated the performance of the model using meteorological traced past record of MHPS generation and consumed electrical power obtained from Sichuan Province in China. The simulation indicated that the technical and financial (income) implications of MHPS may be obtained on the Pareto front by fully considering the performance strategy and target. In addition, the similarity increase in generated power and load is about 7.89% under the proposed energy storage system [33]. Rajadurai et al. proposed SPV-micro for household application that involved an approach of replacing traditional storage equipment. This system operates the DC motor for pumping water to the water storage tank through the pipeline. When power is required, water is released directly to the sump into the MHT, which drives the generator. The authors have outlined the benefits of the proposed system, that is, inexhaustibility, eco-friendliness, and natural regeneration [34]. Syahputra and Soesant proposed micro-hydro/solar photovoltaic amalgamated system for rustic areas in Yogyakarta, Indonesia. The system was optimized using HOMER software and extended particle swarm optimization technique. Based on the optimization results, both micro-hydro from irrigation channel and solar PV energy sources have high potential for generating electricity. The system can generate power that meets energy requirements with surplus [35]. Poudel et al. developed stochastic models for economizing streamflow and empirical model of MHPS in Nepal. The models are integrated in the structure of hybrid 2 tool to simulate the long-term hourly operation of hybrid systems. The results have indicated that MHPS connected with half quantity of batteries in combination with a small pond of one-hour capacity of storing water is capable of supplying electricity sufficiently in compliance with stipulated energy by the Tier 3 level under tracking frame [36]. Das et al. designed the micro-hydro/solar PV/battery hybrid system to solve energy shortage in southern areas of Bangladesh. The system was evaluated by considering the potential of tidal energy at the Bay of Bengal. System optimization for technical and economic feasibility analysis was done using HOMER software. The cost-effectiveness of the system was dependent on location. In addition, comparison results between the proposed system versus diesel generator indicated that total NPC of the former is less by 27% and the estimated COE is about 0.21 $/kWh. The COE for the hybrid is less than that of DG by 28% [37]. Clements et al. carried out the feasibility analysis of hydropower mini-grids for preparing meals using electricity as another alternative to

direct biomass-based cooking in ten families in Simli in rural community, Western Nepal. Data on cooking record and electric mini-grids were gathered, and their compatibility with the micro-hydropower grids and cooking practices is analysed. It was found that the operational cost for cooking using electricity was less expensive than the conventional wood-based cooking method. It was also suggested that the use of electricity for cooking in rural areas of Nepal is not affordable [38]. Hoseinzadeh et al. hybridized hydropower plant with solar PV and wind system for solving energy shortage of the power during the dry season in Iran. Mashhad Sarrood hydropower plant of 61 kW was used in the study. The optimal configuration was obtained using HOMER software. This was done by considering the fluctuations of the river flow and solar radiation. The findings indicated that the energy dearth improvement should be carried using suitable financial appraisal [39]. Das et al. optimized solar PV/biogas/pumped hydro-energy storage/battery bank hybrid system for radio transmitter station in India. The system performance was analysed using meta-heuristic optimization techniques such as water cycle algorithm and month of flame. The results of the aforementioned optimizations were compared with the findings obtained using the genetic method. The study intended to abate the NPC under the defined constraints. Sensitivity investigation was employed to evaluate the reliability of the system using loss of load probability. The study indicated that both suggested meta-heuristic approaches are suitable for acquiring finest optimal design. The optimized system configuration by water cycle algorithm obtained solar PV panel area equals 548.7 m^2 producing power capacity of 69.2 kW, biogas gas generator of 16 kW, battery banks of 21 units, and upper tank volume of 2081.5 cubic metres. The total NPC for the optimized system is $ 0.813 mil. The system is cost-effective [40]. Patel and Chowdhury carried out the techno-economic feasibility analysis of micro-hydro with battery energy storage micro-grid system as a solution to rural electrification for rural areas in South Africa. HOMER Pro was used to optimize the system. The strength of the micro-hydro-resource on the price of electricity was examined. The findings of this study indicated that the stronger the resources the less expensive the system [41]. Zema et al. carried out the techno-economic feasibility analysis of micro-hydropower plants within irrigation systems in Calabria, Italy. The study proposed a simplified method of siting turbines by choosing their power and evaluating the charges and gains. Also, they provided suitable indicators for the design of MHPS in the existing irrigation systems. The results, through the case study, have highlighted the tiniest cost-effective turbine with power capacity of 5 kilowatts. High potential of MHPS from the existing system has much more gains on monetary investment than Italian market. More profits are particularly obtained during the rainy season than in the dry season in the average of 55% [42]. Halabi and co-researchers examined two distributed electric power stations in the area of Sabah in Malaysia. The arrangement of every station included dissimilar components such as solar PV, DG, power-converting devices, and battery banks. The authors confirmed the data, simulated various scenarios such as decentralized DG, PV-DG battery, and PV battery for both stations, and compared the optimal designs based on the selected scenarios. They studied the performance according to the different infiltration levels to find out the effect of PV integration. The investigation was carried out using HOMER software. Technical, economic, and environmental viability was the criteria for comparison of the scenarios. The energy demand was sufficiently met at the least NPC and LCOE. Generally, the findings of the study indicated that the growth trend of energy production using sources of renewable energy was higher than standalone DG. In particular, PV-DG battery is more practically, economically, and ecologically sustainable than DG. In addition, PV battery was found to be the most nonpollutant system with greatest cost. It was observed that the existence of non-conventional power sources improves the operation of decentralized power systems and minimizes the requirement for energy storage mechanisms [43]. Hossain and his co-authors carried out the feasibility analysis of hybrid renewable energy to supply power to big tourist centre located in the South China Ocean in Malaysia. The area receives electric power supply from the decentralized DG. HOMER software was used as a research platform to perform the techno-economic analysis. The highest and average daily electrical load demand for the tourist centre was found to be 1,185 and 13, 048 kilowatts, respectively. The finest optimal decentralized hybridized renewable energy-based system is build up by components such as solar PV, wind, DG, converter, and battery bank. The optimization results of NPC, LCOE, renewable fraction, and emissions of carbon dioxide (CO_2) were found to be $ 17.5, $ 0.279/kWh, 41.6%, and 2, 571, 131 kg/year, respectively. The optimal results of NPC, LCOE, and DG system only were $ 21.09 mil, $ 0.343, and 5, 432, 244 Kg/year, respectively. The proposed hybrid renewable energy-based system is economically and environmentally better than decentralized DG system only [44]. Olatomiwa plus his fellow researchers studied the potentials of the solar and wind in the geographical and political areas of Nigeria. In the study, researchers used the meteorological data in the range starting from eighteen to thirty nine years. Also, the authors analysed the financial and technical viability evaluation of multiple energy source system, which includes solar PV, wind turbines, DG, and batteries. The system was anticipated to provide the electric power to the healthcare building and ensuring that the demand is met. HOMER software was used for carrying out the investigation. The observations of this study indicated that the Sokoto and Joe regions are located in high potential of wind energy, while all the six regions receive the sufficient levels of solar energy for various applications. Optimal results have indicated that the hybrid energy system is the finest choice for the all places. The first configuration of optimal hybrid energy is made up of solar PV-wind-DG batteries that are considered for supplying electricity to rural health centre applications at Enugu, Maiduguri, Sokoto, and Jos. The second one of the hybrid energy system is built by solar PV-DG batteries that have been considered the perfect option for supplying electricity to health facilities in distant places of Iseyin and Port Harcourt because of the abundance of

renewable energy sources. The third configuration of DG system only has highest LCOE and CO_2 emissions of $ 0.911/unit (kWh) and 9211 Kg/year, respectively, in all places under study [45]. Olatomiwa implemented the analysis of decentralized hybrid renewable energy-based systems to supply electricity to the health facilities in the off-grid rural areas or in the areas receiving unreliable power from the grid in Nigeria. These hybrid energy systems intended to address the problem of excessive motherly and child death rates in the country. The researchers described the statistical investigation of solar and wind energy source potentials for some particular remote areas according to the long period of time (in hours) and daily climatological information. The potentials of the energy sources have been validated by the techno-economic analysis of designing and sizing the solar PV-wind battery using HOMER software. The findings of this research indicate that Sokoto and Jos are located in high wind regimes for large power wind systems, while the rest sites are appropriate for small wind applications. On the other side, all sites receive sufficient coverage of solar energy for solar applications. The optimized solar PV-wind-DG battery hybrid system for rural health care is suitable for Sokoto, Iseyin, Maiduguri, Enugu, and Jos. The optimal PV-diesel battery hybrid system is suitable for Port Harcourt due to the available high potential of renewable energy sources. The authors concluded that the high potential of wind and solar sources in the nation makes a best situation for including the renewable energy sources in the planning and execution of the decentralized energy systems in the country for improving the provision of rural health care [46].

The novelty of this study is to explore the feasibility of generating electricity using abandoned potential of nonperennial tributary particularly in the areas of prolonged period of high rainfall (installation of small ROR power station hybridized with solar energy plus batteries). Also, this study presents the utilization of traditional approach of measurements to acquire the input parameters of the tributary. The objective of this study was to optimize the renewable energy-based hybrid system to generate electricity for the selected site. This study explores the neglected hydro-energy potential of a nonperennial tributary incorporated with the availability of abundant solar energy source. The tributary is deemed to have considerable untapped high potential of generating electricity for rural communities in the regions of prolonged high rainfall of Tanzania. This research work tries to explore more diversification of local renewable energy sources. In the study, the estimated water potential derives from abandoned potential of existing tributaries in the country including the Ikata tributary, which was designated as a case study for this research. The tributary is elevated and has suitable water head and sufficient water flow during the rainy season that may be adequate for supporting the rural electrification in the nation. Unfortunately, this potential of the tributary remains untapped. In the rainy season, this tributary has large amount of water low, which is simply flowing down into the permanent river called the River Shongo. This river provides water mainly for drinking and small agricultural irrigation in the community. The study is anticipated to create awareness about how the kinetic energy of water from the disused potential of nonperennial tributary in zones of plentiful precipitation and solar energy source through the hybridization of the two renewable energy sources and battery energy can be harnessed. The study has also taken into account the benefit of high solar radiation availability. Tanzania is located in proximity to the equator, thus receiving substantial amount of solar energy. The potential of water source from a nonperennial tributary and solar energy source is therefore combined into hybrid power generation system. The hybrid system is optimized by the minimization of the costs such as total NPC and COE.

Secondly, in many feasibility studies that have been carried out, there is a limited use of traditional methods for data gathering on hydro-energy source. Traditional methods for the measurement of parameters of water flow rate and head refer to methods in contrary to the advanced ones. They are manually operated and use simple materials and tools for measuring the aforementioned parameters. Traditional methods are bucket methods, float method, and velocity area method, while approaches used for measuring head are, for example, dumpy levels and theodolite. These methods are simple and use simple local materials and inexpensive measuring tools. However, traditional methods need more number of persons to perform the task for measurements and can get hindrance during bad conditions such as massive rain and floods. They have also limitations in terms of accuracy and suitability for only small magnitude of water flow rate [47, 48]. These traditional methods may be supportive in developing countries where poor economy is a major challenge. In this research work, the techno-economic feasibility analysis is performed for increasing the access to electricity to small communities located at Ikukwa Village in Mbeya Rural District, Tanzania. This place consists of nonperennial tributary called Ikata, and its water flow is running down from the hills surrounding the peak of Mount Mbeya. In this feasibility study, case study is carried out at Ikata tributary using its nonperennial hydro-energy source to optimize hybrid micro-hydro-solar PV. Hybrid Optimization of Multiple Energy Resources (HOMER) Pro software is used for the techno-economic analysis for obtaining the optimal configuration of the proposed hybrid system. The analysis of this study has been presented in two stages. In the first stage, the Hybrid Optimization of Multiple Energy Resources (HOMER) software has been used to define the best possible configuration at the lowest total NPC plus COE.

Nowadays, many researches are being carried out using various artificial intelligent (AI) methods. Among these, AI methods include the newly invented artificial intelligent (AI) optimization approaches called PSO, GWO, and GWO-PSOHD. The major merits of the PSO are uncomplicatedness, easiness of procedure, great merging degree, less requirement of storage, and reduced reliance on the preliminary points. The GWO algorithm consists of features such as great convergence degree, performance independence, and heftiness [49]. The GWO-PSOHD optimization method is also lately established of which GWO is combined with PSO for improving the working ability of GWO. The GWO-PSOHD technique has been verified of being simple,

effective, easy, and competitive. Therefore, by employing the above named AI approaches further optimization results have been found out. The AI algorithms have been implemented in the MATLAB program. Through the optimization results, which are obtained from the HOMER software and AI methods, comparative study has been made for validation purposes. The AI optimization techniques provide lower values of total NPC and COE of the proposed hybrid energy system. In summary, this study contributes to the further exploration of unexploited potential in order to diversify renewable energy sources for rural electrification. Moreover, in the paper the estimated capacity of the tributary under study has been estimated through the traditional approach, which is rarely used. Also, the paper involves applications of AI techniques such as PSO, GWO, and GWO-PSOHD algorithms for sizing off-grid system, which further reduces NPC and COE. Similarly, in the paper the system is optimized by taking into account the system reliability.

The structure of this study is as follows: generating electricity through hybrid renewable energy-based system is presented in Section 2. Energy data collection and analysis are provided in Section 3. Electrical load demand is explained in Section 4. The proposed hybrid renewable energy-based system is described in Section 5. The results for technical and economic analysis obtained by HOMER software is provided in Section 6. PSO, GWO, and GWO-PSOHD methods (AI optimization algorithms) are explained in Section 7. Optimization results of technical and economic analysis obtained by GWO technique are delineated in Section 7. Recommendation for the long-term sustainability of the tributary under study is presented in Section 8. Future research work is offered in Section 10.

2. General Description of Electricity Generation through Hybrid Renewable Energy-Based System

There is abundance of electricity in Tanzania though this high potential has not been fully utilized to meet electrical demand, particularly in the rural and isolated areas. Table 1 indicates Tanzania's potential of renewable energy sources [5]. Table 1 indicates Tanzania's potential of renewable energy sources.

Nowadays, the URT immensely supports renewable energy resources for electricity generations. It has been projected that 100% of the total generated electricity in the country in the year of 2050 will be produced from renewable energy sources. In this projection, total power capacity will be approximately 60 gig watts [5]. Hydro- and solar energy sources are not only technologies but also preferable. Power generated from hydro-energy source is reliable and affordable. Solar energy-based systems have been installed in homes, health centres, dispensaries, and schools. Descriptions for hydro- and solar energy sources are given as follows.

2.1. Hydro-Energy Resource. Generating electricity through hybrid renewable energy system is a combination of different but complementary sources of energy. Hybrid technology is reliable and economical. Hybrid power generation configuration is basically subdivided into three categories such as series, parallel, and switched hybrid configuration. Of these three categories, parallel hybrid configuration is the most preferable because it is reliable and energy source within it operates independently. Micro-hydro- and solar PV electric systems are dirt-free systems. Micro-hydro-generating system changes the dynamic energy of water into mechanical power by means of turbine, which is then converted to electricity by generator. Micro-hydropower (MHPS) according to the generating capacity refers to system with a range of capacities starting from 5 kW to 100 kW. Besides, the capacity for MHPS may reach up to 200 kW. Power output generated by MHPS depends on water head and flow rate. This power can be estimated as follows [28]:

$$P = \rho \times Q \times g \times H, \tag{1}$$

where P = power output produced by MHPS in watts (kW), ρ = water density in Kg/m3, Q = water flow rate in m3/s, and H = water head in m. MHSP can generate power for enhancing socioeconomic activities in the rural societies.

Most of the hydropower plants are classified ROR hydropower plants. The power production of ROR plants is not fixed and naturally fluctuates daily, weekly, and seasonally in river discharge. ROR plants can only generate electricity if sufficient amount of water flow of the river is available. If the river discharges fall below the minimum technical inflows, electricity generation ceases. It is worth noting that ROR hydropower projects with pondage as opposed to those with pondage can store water for greatest power consumption and continuously for base load, especially during the rainy periods. Advantages of the ROR power plants without pondage/reservoir for energy production are less impact on the available ecosystem; design is less complicated; overall lower capital costs; smaller water heads; limited time of construction; and offers the possibility of decentralized electrification, relatively low operational and maintenance cost. The design and operation are a challenging task in terms of robustness, flexibility, and periodic fluctuations.

2.2. Solar Energy Resource. Tanzania is a tropical country with a high potential of solar energy within a range of 4–7 kWh/m2/day [25]. Regardless of this sufficient amount of solar energy, there is a limited installed capacity of decentralized systems in rural areas. A solar energy source is an intermittent in nature causing power output variability. Power output depends on the amount of solar radiation, orientation, temperature, and site conditions [47, 50]. A battery energy storage system is required to store energy during high solar radiation and supply electrical load whenever solar radiation is poor. In a hybrid renewable power generation system, energy storage cost is reduced and its performance is also reliable [47, 48]. Electricity generation by solar PV technology consists of solar panels, charge controllers, deep cycle

Table 1: Tanzania's potential of renewable energy sources.

Renewable energy source	Approximated potential	Growth in percentage
Mega hydro-energy	4.7 GW	12%
Small hydro-energy	485 GW	2%
Solar energy	Greater capacity than 200 W_p/m^2	Solar radiation: 215 W/m^2
Biomass energy	Greater capacity than 500 MW	34 million m^3/year
Geothermal energy	5000 MW	—
Biofuels	Existence of plants for ethanol and biodiesel production	—
Wind energy	Greater speed than 8 m/s	—
Tidal and wave energy	Ongoing assessment	—

batteries, and power converters. Power output is calculated as follows [28]:

$$P_{\text{Peak}} = \frac{E_{el} \times I_{STC}}{E_{\text{global}} \times PF}, \quad (2)$$

where P_{peak} = peak power (KW p) under standard test conditions (STCs), E_{el} = actual annual energy output (kWh/annual), I_{STC} = irradiance solar radiation (1 kWh/m^2) under STC, E_{global} = annual global solar radiation (kWh/m^2/year), and PF = performance factor.

3. Energy Data Collection and Analysis

Based on the novelty and proposed hybrid MHPS-solar PV, a case study is made at the location where there is a high potential for a nonperennial tributary in a prolonged rainfall including the high potential of solar energy. The following sections provide a brief description of area of study and methodology of this study.

3.1. Description of Area of Study. Ikukwa Village is one of the villages in Mbeya Rural District of Mbeya Region (latitude 8° 42.6′ S, longitude 33°27.9′E), Tanzania. The village comprises 14 sub-villages with 570 total households. Geographically, and logistically, the village is located beyond Mount Mbeya where it has been difficult for direct grid extension from Mbeya City. The location of this study is latitude 8°45.4′ south and 33°17.4′ east, as indicated in Figure 1. This research is carried out in three main phases such as determination of theoretical capacities of solar photovoltaic system (SPVS) and MHPS, the analysis of energy demand of Ikukwa Village, and the optimization of proposed hybrid system. Hybrid optimization multiple electric resources (HOMER) Pro software is used for simulation during the techno-economic analysis of the hybrid energy micro-grid system.

Logistically, during the rainy season the road to the area of study is also impassable. Recently, only 40 households of the total number of households of the village are connected to the national grid distributed by Tanzania Electricity Supply Company (TANESCO). The availability of grid connectivity in the whole village is equivalent to 7%. One of the reasons is that this village has dispersed households and poor population density. Table 2 indicates the status of grid connectivity percentages by sub-villages. This village consists of tributary called Ikata. The Ikata tributary is a



Figure 1: Location of Ikukwa Village in rural areas Mbeya Region, Tanzania.

Table 2: Status of national grid connectivity (TANESCO) at Ikukwa Village (source: survey by authors before 31.07.2019).

Subvillage	NH	Grid-connected customers	% of grid-connected NH
Mdonya	26	0	0
Kiwanja	38	0	0
Unguja	28	0	0
Shongo	33	01	0.2
Juakali	42	0	0
Ujamaa	34	01	0.2
Kariakoo	41	0	0
Ukwaheri	29	0	0
Ikukwa	52	14	2.5
Mbuwi	42	0	0
Zongo	59	08	1.4
Itende kati	90	10	1.8
Mahonza	28	05	0.9
Itende juu	28	01	0.2
Total	570	40	Approx.7. %

NH = number of households.

nonperennial water stream having a high potential of generating electric power during a rainy season. This stream of water flows down from the hills surrounding Mount Mbeya to the River Shongo. Figure 2 indicates the Shongo River, which receives water supply from Ikata tributary at the village. Other tributaries that deliver water supply to River Shongo are Ho, Ilwilo and Igodela. The perennial River Shongo supports Ikukwa dwellers by providing them with water supply for drinking and irrigation for agricultural activities by means of small channels. According to the fieldwork carried out by the authors, Ikata tributary is more



Figure 2: Shongo River receives water supply from Ikata tributary at Ikukwa Village from the hills of Mt. Mbeya, Mbeya Region, Tanzania (photo during the dry season).

preferable for distributed electric generation than the Shongo River because the former has a substantial water head and is closer to the proposed site of study. It is well known that distributed generation is installed close to the electrical load for improved reliability and efficiency. For these reasons, Ikata tributary is considered to have a high potential of generating electricity when hybridized with solar PV-generating system. Secondly, Mbeya Region where Ikata tributary is found generally has high rainfall for a prolonged period of time for sufficient water flow rate during the rainy season. The hybrid of MHPS of run-over-river category and solar PV-generating system is intended to increase power accessibility to sub-villages of Ikukwa Village, useful for reducing the challenge of frequent and longstanding load shedding from the national grid network, and can support productive activities in the Ikukwa community.

The steps for this research activity are presented in Figure 3. The steps involved are as follows: this research process follows the steps such as literature review, field appraisal, and energy assessment for the proposed hybrid system including gathering data for miscellaneous meteorological information, energy resource, and demand assessments, and determination of the optimal hybrid renewable energy systems based on the lowest total NPC and COE, judgement is made about whether the analysis gives optimal results or not. If yes, optimal results are presented, and finally, conclusions are made. In contrast, the analysis is repeated until optimal results are achieved.

3.2. Methodology. Collected data for hydro-potential of MHPS were achieved from a field appraisal in the proposed site. The hydro-energy assessment is concentrated on the tributary of Ikata at Ikukwa Village, Mbeya Rural District, Mbeya Region, Tanzania. In this tributary, direct measurements for the estimation of water flow rate and head parameters were obtained through traditional methods. Moreover, the potential for solar radiation and other meteorological information was achieved from the databank of the National Aeronautics and Space Administration (NASA) [51, 52] and Tanzania Meteorological Authority (TMA) via Mbeya Meteorological Station (MMS), respectively. Miscellaneous meteorological data are important for understanding how the changes in temperature and rainfall affect the performance of MHPS and the proposed hybrid energy system in terms of power output. Table 2 indicates the status of national grid connectivity (TANESCO) at Ikukwa Village (source: survey by authors before 31.07.2019) (Table 3).

3.3. Platform Used. HOMER software is micro-grid modelling software. It was established by National Renewable Energy Lab (NREL) under the department of energy in the United States (USA). HOMER is popular leading software used for the optimization of micro-grid system. It is commonly used for feasibility studies to determine the optimal configuration of the designed system. The factor that is used for the selection of the optimal configuration in the simulation is the total lowest net present cost (NPC) and cost of energy (COE) [53].

In addition to HOMER software, the GWO algorithm is used for the technical and economic analysis of the system by incorporating the system reliability constraint. In this specific study, loss of electrical power probability (LEPP) has been taken into account as reliability restriction. The optimum system is obtained at the lowest total NPC (also the lowest COE) under reliability boundaries. The detailed description of the technical and economic parameters used in the analysis is as follows [54]:

$$LEPP = \frac{\sum_{t=1}^{T} SPS(t)}{\sum_{t=1}^{T} P_L(t)}, \tag{3}$$

where SPS is the shortage in power supply, P_{L} is the total electrical load, and T is a period of time for a year, which is equivalent to 8760 hours. If LEPP equals zero, it implies that the hybrid micro-grid system produces sufficient power for the load, and if the LEPP is 1, it means that the system is incapable of satisfying the load [40].

$$SPS = P_L - \left(P_{\mathrm{ghybrid}} + SOC_t - SOC_{\mathrm{min}\,i}\right), \tag{4}$$

where P_{ghybrid} is the total power generated from the hybrid micro-grid system, SOC_t is the state of charge of the battery at time t, and SOC_{mini} is the minimum state of charge of the battery.

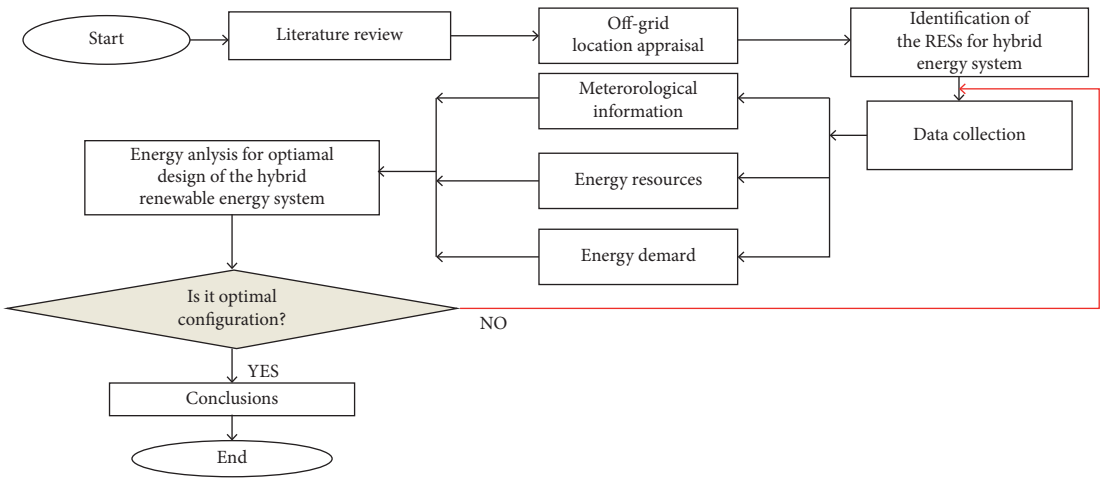


Figure 3: Steps for implementing the research.

Table 3: Total electrical load at Ikukwa Village (7% access to grid from TANESCO) (source: survey by authors before 31.07.2019).

Name of small communities of Ikukwa Village	Estimated load in watts (W)
Ujamaa	115 803
Ikukwa	149 531
Jua kali	141 482
Kiwanja	52 306
Ukwaheri	46 079
Itende juu	33 063
Mbuwi	80 626
Mahonza	49 799
Kariakoo	79 475
Unguja	33 404
Mdonya	61 372
Zongo	100 825
Shongo	42 448
Itende kati	129 953
Total	1116166

TANC comprises all annualized costs of the system in $/year. It consists of annualized investment charge (AIC), annualized spare charge (ASC), annualized fuel charge (AFC) of energy source, and annualized operation and repair charge (AOMC) of the hybrid system. TANC is computed in the following equation [40]:

$$TANC = ACC\,(MHPS + \text{Solar}\,PV + \text{Baterries} + \text{Converter}) + ASC\,(MHPS + \text{Solar}\,PV + \text{Baterries} + \text{Converter}) + AFC\,(MHPS + \text{Solar}\,PV) + AOMC\,(MHPS + \text{Solar}\,PV + \text{Baterries} + \text{Converter}). \quad (5)$$

However, in this study AFC is neglected as hydro- and solar energy sources are free of charge, and for simplicity, salvage value is not considered.

The NPC can be simply defined as a sum of whole expenses of project's lifespan as follows [40]:

$$NPC = \frac{TANC}{CRF\,(i,\ Pproj)}, \quad (6)$$

where i indicates the interest rate in %, Pproj is the lifespan of the project in years, and CRF represents the capital recovery factor.

COE ($/kWh) refers to the average charge per unit of useful generated electricity. COE can be computed as follows [40]:

$$COE = \frac{(NPC * CRF)}{EL}, \quad (7)$$

where EL is the total annual energy consumption (kWh/year). COE is alternatively computed in terms of TANC by the following equation [40]:

$$COE = \frac{TANC}{EL}. \quad (8)$$

CRF is defined as follows [40]:

$$CRF = \frac{i_{r*}(1+i)^N}{(1+i)^N - 1}, \quad (9)$$

where CRF is defined as the capital recovery factor, i_r equals the interest rate, and N is the lifespan of a hybrid micro-hydro/solar photovoltaic system with battery energy storage. i_r is defined as follows [40]:

$$i_r = \frac{(\text{i} - \text{inflation rate})}{(1 + \text{inflation rate})}. \quad (10)$$

4. Assessment of Energy Resources, Meteorological Information, and Electrical Load Demand

The main goal of this research is to optimize the hybrid for the same mentioned technologies. A hydro-energy source originates from a nonperennial Ikata tributary. Hydropower technology requires two major parameters for power generation to meet forecasted demand. These required parameters are suitable to head and water flow rate. The assessment of the parameters needs deep analysis of a number of facts regarding water stream such as rainfall, nature of ground, and vegetation at certain period of time. There are a number of methods, and both traditional and modern methods are employed for determining the abovementioned parameters. However, traditional methods are useful for the initial estimation of energy potential, especially in a limited time, and are mostly applicable in developing countries. Therefore, data for water head and water flow rate were collected by traditional methods. Data for solar radiation were obtained from the NASA-SSE, and energy demand was collected from the inventory of appliances.

4.1. Assessment of Hydro-Energy Potential of Ikata Tributary. Part (a): assessment of hydro-energy resource for MHPS using the traditional approach.

Water flow rate and head are required input parameters to HOMER Pro software and are determined using traditional methods as follows [55].

4.1.1. Materials and Methods for the Determination of Water Flow Rate. The estimation of water flow rate float method is used. This method requires the determination of surface water velocity, width, and depth of water stream. Finally, these dimensions enable the calculation of water flow rate.

(1) Estimation of Average Velocity of Water Stream. Cylindrical wooden coloured floats of individual diameter and length of 5 cm and 30 cm, respectively, are used to assess water surface velocity of River Ikata. The distance of 10 cm between two marks by means of two pieces of bamboo tree was measured using a measuring tape. Stopwatch was used to monitor time in seconds for float movement over marked distance. Ten-time measurements were recorded and lastly averaged. Table 4 shows time measurements.

Sum = 16 + 16+16 + 15+16 + 15 +16 + 17 +16 + 16 = 159.

Therefore, the average time for the movement of float at the measured distance is 15.9 s. Surface water speed equals the distance along the water stream per average time, which is 10 m/15.9 = 0.63 m/s.

Average water velocity V = surface water speed × correction factor = 0.63 × 0.85 = 0.536 m/s.

(2) Estimation of Average Width of Water Stream. The average depth of water stream was measured using the traditional method by means of pieces of bamboo tree and measuring tape. Pieces of bamboo tree were laid at equal spaces across water stream, and the distance between two ends was marked. At the end of this process, pieces of bamboo trees were taken aside, and from two marked ends of individual piece of bamboo tree, distance, which is equivalent to water stream, was measured by a measuring tape. Table 5 indicates width measurements.

As 6.1 m has the highest frequency among other width measurements, therefore this value is used as the average width of water stream of Ikata River. Therefore, $W = 6.1$ m.

(3) Estimation of Average Depth of Water Stream. The depth of water stream was measured using pieces of bamboo tree and measuring tape. Several pieces of bamboo tree were dipped vertically to the bottom of the water stream along the width of water stream, and the water level of each bamboo stick was marked using marker pens. These sticks/rods were taken out for measuring distance starting from the marked point of individual piece to the bottom. Table 6 shows depth measurements.

The deepest measurement from the table above is 1.8 m. The estimation for average depth of water stream is half of the deepest measurement; that is, $d = 1.8/2 = 0.9$.

(4) Determination of Water Flow Rate. Having estimated average water velocity, average width and average width water flow rate can be found. A water flow rate of the river is the product of three above named parameter. Therefore, water flow rate = $\text{V} \times w \times d = 0.536\ \text{m/s} \times 6.1\ m \times 1.3\ m = 2.943\ \text{m}^3/\text{s}$.

4.1.2. Materials and Methods for the Determination of Water Head. Water head is determined through traditional or local methods known as the carpentry method. Materials and equipment required are measuring tape, vertically erected wooden piece, that is, pieces of bamboo tree (naturally marked), and horizontal rectangular board and mason's level. A piece of chalk or marker pen has been used for marking the height of the bamboo during the measurement of water head using the carpentry method. Measuring measurements are used to take measurements for distance; pieces of bamboo are vertically erected to measure height; horizontal board and mason's level are used for horizontal levelling from reference on the ground to required height on a vertical piece of bamboo (marked by chalk or marker pen).

Table 4: Time measurements.

Measurement	Time (s)
1	16
2	16
3	16
4	15
5	16
6	15
7	16
8	17
9	16
10	16

Table 5: Width measurements.

Measurement	Width (m)
1	6.7
2	6.1
3	6.1
4	6.1
5	7.0
6	6.8
7	6.9
8	6.6
9	6.0
10	5.6

Table 6: Depth measurements.

Measurement	Depth (m)
1	1.6
2	1.6
3	1.5
4	1.3
5	1.7
6	1.8
7	1.4
8	1.2
9	1.4
10	1.3

Before estimating water head, intake position and location of turbine are identified. Table 7 displays water head measurements.

4.1.3. Calculation of Hydropower Capacity. Theoretical power potential depends on two major factors, namely flow rate and head of the falling water. The potential of micro-energy resource is evaluated using (1) as follows:

$$\begin{aligned} \rho &= 1000\,\mathrm{Kg/m^3}, \\ g &= 9.8\mathrm{m/s^2}, \\ Q &= 2.943\mathrm{m^3/s}, \\ H &= 13\mathrm{m}. \end{aligned} \tag{11}$$

Therefore,

Table 7: Water head measurements.

Measurement	1	2	3	4	Gross head in hills (m)
Height (m)	3.4	3.4	4.00	2.6	13

$$\begin{aligned} P &= 1000 \times 9.81 \times 2.943 \times 13 = 375,320.79\ W, \\ P &= 375\,kW. \end{aligned} \tag{12}$$

This is the theoretical maximum power capacity. This capacity is expected in the month of January due to the highest rainfall in the location of the site project.

Part (b): assessment of hydro-energy resource using HOMER Pro software.

Depending on the surveyed parameters of Ikata tributary implemented in HOMER Pro software, Figure 4 indicates the annual monthly hydro-flow rate. The scaled annual average water flow rate is 899 litres per second. The comparison of the two figures, namely Figures 4 and 5, indicates the fluctuation of rainfall in a year and it affects the water flow rate of the tributary under study, hence the power output from MHPS within the proposed hybrid technology. The drastic effect of poor rainfall on water flow in the tributary is overcome by the integration of SPVS with batteries. The focus of this research is to investigate the power generation likelihood of the tributary and the general functioning of the proposed hybrid energy micro-grid system.

4.2. Solar Energy Resource. Data for monthly solar radiation are obtained from the database of NASA [51]. Mbeya Region is located at latitude 8° 42.6′ south and longitude 33°27.9′ east. Data for monthly mean solar global horizontal irradiance in terms of solar radiation and clearness index are presented in Table 8 [51]. Data on solar energy for Mbeya Region are used for a selected study area in Mbeya Rural District. The annual average of solar radiation analysed by HOMER software for a proposed project site is 6.11 kWh/m^2/day. The potential of solar energy for the site and global solar resource are given in Figures 6 and 7, respectively.

4.3. Miscellaneous Meteorological Information. Meteorological information is obtained from Tanzania Meteorological Authority (TMA) at the branch of Mbeya Meteorological Station (MMS) [52]. The required meteorological data for this study are rainfall and temperature. In the month of January, the area of study has the highest average rainfall about 247.78 mm. The average rainfall for May and October is 11.60 mm and 16.70 mm, respectively. In the period from June to September, the rainfall is zero. The amount of rainfall in the months of May and October is low. However, water flow from the tributary under study in such months though is assumed to be sufficient to generate power depending on geological conditions. Rainfall has a large impact directly on the amount of water flow rate of the nonperennial tributary and indirectly on the operation of MHPS in the proposed hybrid energy system. Therefore, no power will be produced from MHPS. Instead, SPVS with battery energy storage will produce power within the hybrid



Figure 4: Annual monthly average water flow rate.

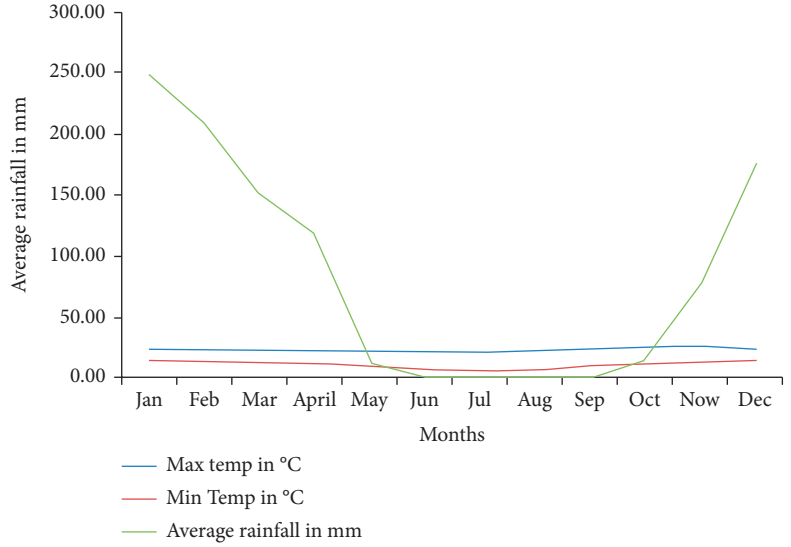


Figure 5: Rainfall and temperature for the location of the proposed project in Mbeya Region, Tanzania (2000–2015) [52].

Table 8: Data for monthly solar radiation and clearness index [51].

Month	Clearness index	Solar energy radiation kWh/m^2/day
January	0.48	5.27
February	0.51	5.54
March	0.55	5.83
April	0.60	5.89
May	0.67	5.96
June	0.71	5.96
July	0.72	6.14
August	0.71	6.57
September	0.69	6.94
October	0.57	6.95
November	0.51	6.21
December	0.61	6.06

energy system from May to September [52]. The period ranging from June to September is now generally considered the dry season. The minimum temperature is experienced in July at around 6.10 °C, while the maximum temperature is in November at about 26.80 °C [52]. Table 9 indicates collected data for meteorological information (2000–2017) [52]. Table 8 and Figure 7 indicate data for rainfall and temperature for the proposed site of study.

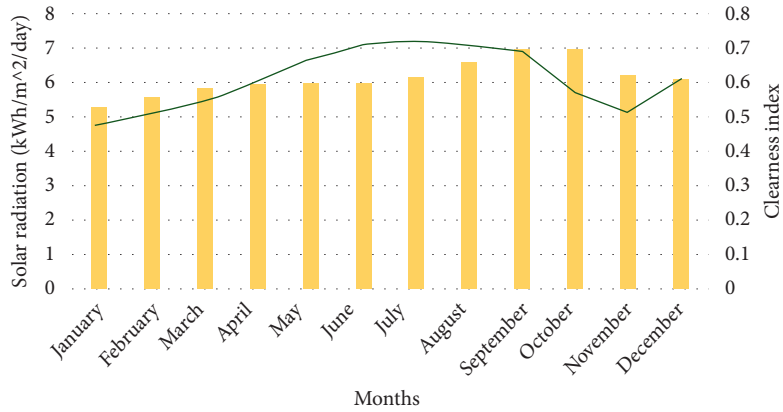


Figure 6: Potential of solar energy for the proposed site.

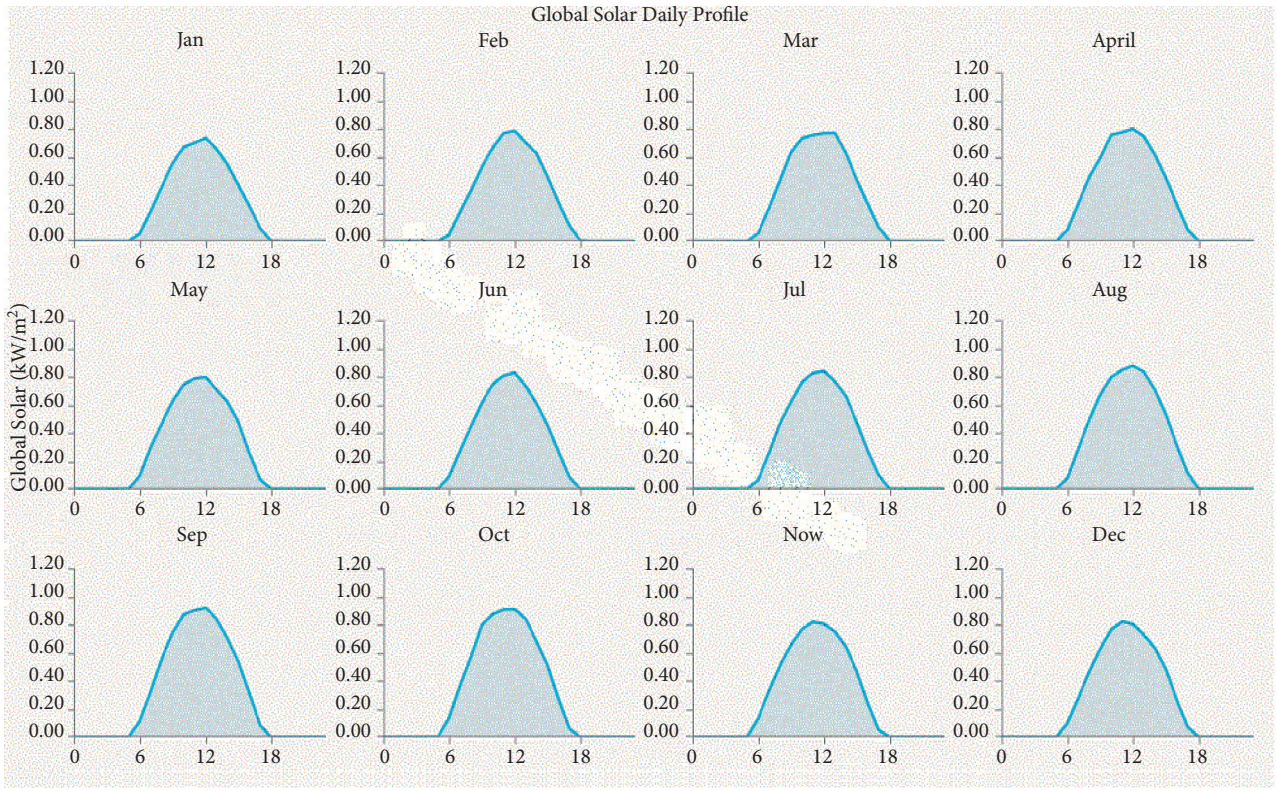


Figure 7: Annual global solar energy resource.

4.4. Electrical Load Demand. Ikukwa Village, Mbeya Rural District, Mbeya Region of Tanzania, consists of 14 small communities. The load demand in this area of study is found in four categories such as residential, institutional, small commercial, and institutional and small-scale industrial loads. Normally, in most of the rural and remote areas electrical load demand is dominated by the residential load. As shown in Table 2, only 7% of all communities have access to grid connectivity, particularly domestic houses. This limited grid coverage is due to scattered population and shortage of budget. This study considers only 7 communities without access to electricity. Other remaining small communities with at least access are not considered on the assumption that all loads are connected to the distribution system of the grid. Therefore, data collection for the electrical energy demand of small communities without electricity for 100% is presented in Table 10. Electrical demand for such communities has been estimated by collecting data on rated power of anticipated appliances. In the first place, these data have been deliberately distributed over twenty-four-hour period on the excel sheet. The maximum estimated daily load profile is around 30 kW. Moreover, similar data for the communities with limited access to electricity have been implemented in HOMER Pro software. Electrical load profiles in tropical areas such as Tanzania are considered the same every month of the year. The reason why is

Table 9: Collected data for meteorological information (2000–2017) [52].

Month	Avg. Rainfall in (mm)	Max. temperature in °C	Min. temperature in °C
January	247.78	24.10	14.70
February	207.78	24.40	14.50
March	150.50	24.40	14.30
April	117.64	23.50	12.80
May	11.60	23.10	10.00
June	0.00	22.30	7.10
July	0.00	22.40	6.10
August	0.00	24.00	7.70
September	0.00	26.00	10.80
October	16.70	26.00	12.30
November	78.10	26.80	14.50
December	175.20	24.90	24.90

Table 10: Total electrical load for small communities at Ikukwa Village without access to electricity (source: survey by authors before 31.07.2019).

Name of the community without electricity	Estimated load in watts (W)
Juakali	98072
Kiwanja	52306
Kariakoo	79475
Mbuwi	80626
Mdonya	70790
Ukwaheri	46420
Unguja	33 404
Total	461093

that there is no extreme weather variation. The daily random variability has been assumed to be zero. The estimated Ac load (baseline) for the site is as follows: average energy is 511.1 kWh/day, average power capacity is 21.3 kW, peak power is 30.31 kW, and load factor is 0.71. Figures 8–10 indicate the daily electrical load profile for the sub-villages without electricity connectivity from the national grid obtained by excel datasheet, daily electrical load profile simulated by HOMER Pro software, and monthly average seasonal load profile, respectively.

In the first place of this study, the analysis for the proposed system comprises the traditional approach of measurements and soft computing technique (HOMER software). In the absence of measured and financial limitations, traditional methods of measurements have been used to determine the profile for the tributary under study. This traditional approach of measurements has been used to determine the water flow rate and head, which are important parameters for an approximation of expected maximum generated output from the MHPS resulting during the rainy season.

MHPS is amalgamated with solar PV and batteries constituting the proposed hybridized energy system. Then, the traditional approach has been validated in HOMER software to acquire the optimized configuration of the proposed system at the lowest total NPC. This type of software is mostly used for optimization and sizing of the hybrid energy system. Nevertheless, optimization by the aforesaid software has some limitations such as limited multiple objective functions; results of the COE are not arranged correspondingly in alignment with the results of total NPC, no consideration of depth of discharge (DOD), hourly variations of the parameters, and fluctuations of bus electrical energy profile. To deal with the abovementioned drawbacks of using HOMER software for optimization purpose, researchers have been constantly modernizing optimization methods. These advanced optimization techniques are termed as meta-heuristic methods. The grey wolf optimization (GWO) technique is one of the meta-heuristic methods. The merits of GWO approach are simplicity and adaptability [49, 56].

Therefore, in addition to soft computing optimization using HOMER software, the optimization of the total NPC has been carried out using AI techniques such as PSO, GWO, and GWO-PSOHD algorithms for obtaining supplementary optimal results. These AI methods are also used to minimize the total NPC of the proposed hybrid power technology according to equation 29. The optimization of the system using the AI approaches has been executed in MATLAB with specified input parameters.

5. PSO Algorithm

The PSO algorithm is one of the stimulated stochastic population-based optimization approaches established since 1995. The developers of this algorithm are Kennedy and Eberhardt, and its application reflects behavioural aspects of birds' flock. Basically, the PSO application involves the simulation of birds' social behaviour of assembling and adjusting themselves in their location while looking for attractive foodstuffs, defending against other greedy beasts using the reciprocated information and communal sensitive intelligence. The benefits of the PSO approach against other investigative optimization techniques are uncomplicatedness, straightforwardness regarding usage, great merging level, low storage needs, and reduced reliance on the preliminary points [57].

The PSO algorithm consists of both individual familiarity and acquired knowledge from working in the gathering to develop the simplified model that can successfully give solution to the complicated problems. This approach employs several agents (also known as particles), which build up the group going around exploring for the finest

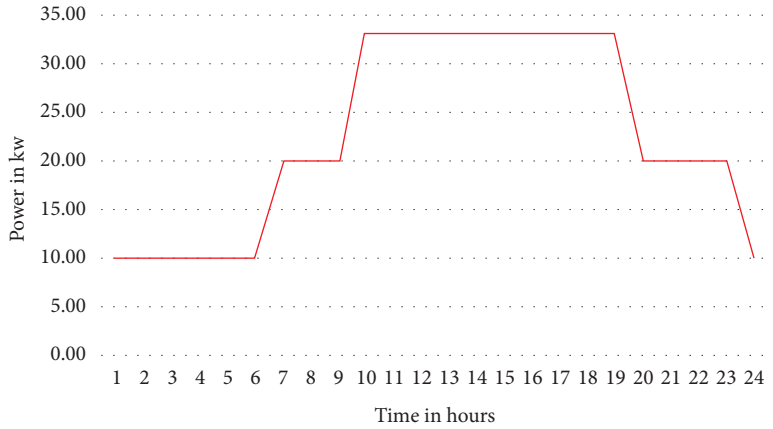


Figure 8: Electrical load profile for small communities without electricity connectivity from the national grid at Ikukwa Village, Tanzania.

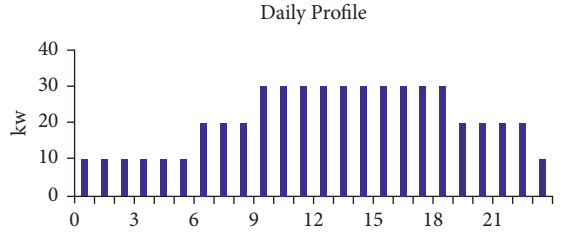


Figure 9: Electrical load over a 24-hour period simulated by HOMER Pro software.

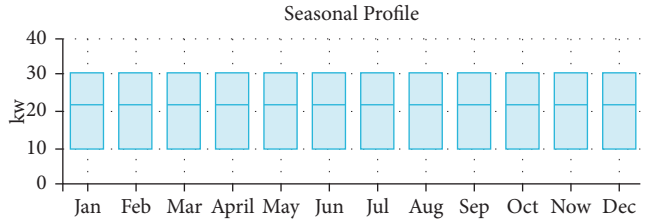


Figure 10: Monthly average seasonal load profile.

answer. Every agent in the exploring cosmos alters its hovering location based on its individual flying knowledge plus the flying practice of other agents in the crowd. Every agent maintains its distinct position in the place of solution. The coordinates are related to its finest fitness rate. This rate is known as personal best (P_{best}). In addition to the P_{best}, there is another finest rate for fitness, which is also taken into account as the best fitness rate in the PSO algorithm, and it is acquired by any agent in proximity to that individual agent. This kind of the finest fitness rate is named as global best (G_{best}) [57].

5.1. Mathematical Representation of PSO Algorithm. In this part, the mathematical model of the PSO approach is defined. The primary notion of the PSO approach is built on the speeding every particle in the direction of its P_{best} plus G_{best} positions at a randomly weighted fast-moving factor at every period of time. Each particle tends to modify its coordinates according to some equation and variables. The present location , speeds, and the distance of the P_{best}, and the distance between the available location of G_{best} are modified. The variation of speed and location of the particle can be represented numerically as follows [57]:

$$\begin{pmatrix} V_{k+1=}^{i} V_k^i + C_1 r_1 \left(p_k^i - x_k^i\right) + C_2 C_2 \left(p_k^i - X_k^i\right) \\ X_{k+1=X_k^i + V_{k+1}^j}^{i} \end{pmatrix}, \quad (13)$$

where V_k^i symbolizes the position of the particle, V_k^i delineates the velocity of the particle, p_k^i stands for the finest recalled location, C_1 and C_2 describe the reasoning and shared variables, and r_1 and r_2 are randomized numbers ranging from 0 to 1.

5.2. The Primary PSO Algorithm. The key steps in the PSO approach are provided as follows [58]:

(i) Initializing the population size of the particles

(ii) Reiterating

(iii) Computing the function's fitness for personal particle

(iv) Adjusting the best particle in the swarm

(v) Picking out the finest particle

(vi) Modernizing the particle's location

Reiterating up to when the required conditions are fulfilled.

Apart from other AI optimization methods, in this research paper the PSO application has been employed to assess the monetary metrics such as total NPC and its parallel COE based on (6) and (7). The PSO algorithm has been executed with a fixed mass of inertia, and input parameters are described in Table 11. The location of a particle inside the search space embodies the power capacities of apparatuses of a given structure of energy system. For a single generation, this PSO method regulates the location of the particles in the swarm, tending to move in the direction of the optimum solution. Every particle in the swarm has the capability of understanding its existing location and speed, that is, its own finest location and the finest location of whole group for a single generation. Deployment of the input technical specifications and power capacity equivalent to the location of an individual particle is estimated for a period of time of one year. Then, in case all specified limits are met, either the fitness of energy system or the total NPC is computed using input economic metrics. Violation regarding any limit by energy system configuration is instantly rejected. Practically, NPC is adjusted to the huge quantity so as to limit the group from being pulled to this location inside the exploration space. The outcomes of the preceding generation are inputted to the succeeding generation. The procedure of the PSO method is repetitive until the greatest quantity of generations is attained. Finally, the solution for NPC is acquired at optimal power capacity [58].

Table 11: Input parameters of PSO approach.

Input variables	Worth
Least search space	0
Greatest search space	10 000
Quantity of iterations	100
Size of population	50
Quantity of generations	200
Weightiness of inertia	0.6
Weightiness of cognition	1.6
Weightiness of community	1.6

6. GWO Algorithm

In the second phase of this study, further optimization results for the proposed energy system are obtained from the GWO technique. The GWO procedure for the first time was suggested by Mirjalili and other researchers in 2014. The key motivation of this approach is the management and pursuing technique of grey wolves. In the mathematical expression, the first group that is known as alpha (a) wolves has been reflected as the best fitting solution. Consequently, beta ($ß$) and delta (δ) are the second and third finest solutions, respectively. The remaining fourth candidate solution is known as omega (ω) wolves. α, $ß$, and d wolves lead to hunting in the GWO optimization, while the group of ω wolves follows the first three packs during searching for the optimum global solution. The societal pyramid of all groups of wolves is illustrated in Figure 11.

Moreover, the following equations are presented for enclosing and chasing the wolves [49, 56, 58]:

$$\overrightarrow{D} = |\overrightarrow{C}.\overrightarrow{Xp}(t) - \overrightarrow{X}(t)|, \tag{14}$$

$$\overrightarrow{X(t+1)} = \overrightarrow{Xp}(t) - \overrightarrow{A}.\overrightarrow{D}. \tag{15}$$

In these given equations, $\overrightarrow{Xp}$, $\overrightarrow{X}$, and t define the position vectors of the prey, grey wolf, and current iteration, respectively. Then, coefficient vectors are denoted by $\overrightarrow{A}$ and $\overrightarrow{C}$, and they are estimated using the following equations [49, 56, 58]:

$$\overrightarrow{A} = 2\overrightarrow{a}\overrightarrow{R1} - \overrightarrow{a}, \tag{16}$$

$$\overrightarrow{D} = 2\overrightarrow{R2}. \tag{17}$$

In the above equations, uneven vectors are denoted as $\overrightarrow{R1}$ and $\overrightarrow{R2}$. They exist in the range starting from 0 to 1, while $\overrightarrow{a}$ is diminished from 2 to 0 in the complete iteration.

The wolves may identify the locality of the quarry and surround them directed by α. Yet, the location of the optimum quarry is not recognized in exploration space. Mathematically, the model for pursuing the performance of the wolves is marked as α, $ß$, and d. They signify the grey wolves knowing fully the potential whereabouts of the prey. Therefore, the foremost three finest answers are stored. Then, ω group of the wolves needs to make the revision of their sites more improved than the present ones as defined in the given equations [49, 56, 58]:

$$\begin{aligned}
\overrightarrow{D\alpha} &= |\overrightarrow{C1} \cdot \overrightarrow{Xa-X}|,\\
\overrightarrow{D\beta} &= |\overrightarrow{C2} \cdot \overrightarrow{X\beta} - \overrightarrow{X}|,\\
\overrightarrow{D\delta} &= |\overrightarrow{C3} \cdot \overrightarrow{X\delta} - \overrightarrow{X}|,\\
\overrightarrow{X1} &= \overrightarrow{X\alpha} - \overrightarrow{A1} \cdot (\overrightarrow{D\alpha}),\\
\overrightarrow{X2} &= X_{\beta} - \overrightarrow{A2} \cdot (\overrightarrow{D\beta}),\\
\overrightarrow{X3} &= \overrightarrow{X\delta} - (\overrightarrow{A3} \cdot (\overrightarrow{D\delta})).
\end{aligned} \tag{18}$$

$$\overrightarrow{X}(t+1) = \frac{(\overrightarrow{X1} + \overrightarrow{X2} + \overrightarrow{X3})}{3}.$$

When the quarry halts from moving, the process of hunting the wolves also terminates. Referring to the given

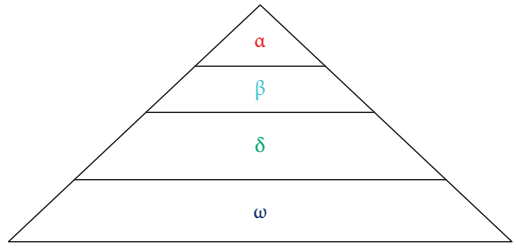


Figure 11: Social ladder of grey wolves.

equations, $\vec{a}$ diminishes from 2 to 0 and therefore A also decreases since $\vec{A}$ is dependent of $\vec{a}$. The quarry is attacked by grey wolves if $\vec{A}$ exists in the range starting from – 1 to 1. Moreover, weights for the definition of distance by the quarry are varied by $\vec{C}$. This value is important for saving answer from the local bests, specifically in the final iterations. General steps executed in the GWO algorithm are summarized as follows [49, 56]:

(i) The establishment of the population of grey wolf X_i ($i = 1, 2, 3\ldots, n$)

(ii) The initialization of a, A, and C

(iii) The computation of the fitness of each pursuit agent, that is, X_α, X_β, and X_δ. X_α, X_β, and X_δ stand for the finest pursuit agent, the 2nd finest pursuit agent, and 3rd finest pursuit agent, respectively

(iv) While ($t <$ Maximum quantity of iterations)

(v) For every pursuit agent

(vi) The modernization of the location of the present pursuit agent

(viii) End for

(ix) The modernization of the α, A, and C

(x) The computation of the fitness of all pursuit agents

(xi) Modernization of the X_α, X_β, and X_δ

(xii) $T = t + 1$

(xii) End while

(xiii) Return X_α

6.1. GWO-PSOHD Algorithm. The GWO-PSOHD approach of optimization is a recently innovated nature-inspired technique. It is one of the hybridized categories of the GWO. GWO is deliberately combined with the PSO method for the refinement of the working capability of the GWO. It is not a scope of this specific study to outline, compare, and contrast the applications of all types of the hybridized forms of GWO. Nevertheless, the hybrid-based GWO algorithm has indicated superior results, specifically convergence level, improved outcomes, effectiveness, and greatest steadiness in more discrete algorithms for *n*. It has the ability of handling both controlled and uncontrolled optimization problems. This method focuses on the optimum solution by holding the improved solutions and terminates the furthermost ones [59, 60].

In this study, a new method called GWO-PS0HD is investigated for the optimization of the total NPC and corresponding levelized COE. Therefore, this study concentrates on the optimization by determining the least total NPC of renewable-based micro-grid of multiple sources.

6.2. Fundamental Practices of GWO-PSOHD Approach. The following are the steps for the GW0-PSOHD approach [59, 60]:

(i) Initializing the search agents and describing the solution space

(ii) Running the GWO method

(iii) Generating the least magnitudes for the complete number of agents

(iv) Permitting the agents to the PSO method

(v) Returning the adapted location back to the GWO method

(vi) Repeating similar procedures up to when the ending conditions are acquired.

7. Modelling of Individual Components Incorporated in the Proposed Hybrid Micro-Grid System

If the total generated power by energy sources integrated into the hybrid micro-grid system equals energy requirement, this means the total generated power meets the electrical load. In case, the generated electricity is greater than the load demand, and surplus electricity is stowed in the batteries (battery charging process). If the energy requirement is greater than the generated electricity, the stored energy is derived from the batteries (discharging process) to meet the energy requirement. In some conditions, energy demand is greater than the sum of generated electricity and stored energy. In such a situation, the portion of the load demand needs to undergo shedding, thus resulting in a loss of electricity supply. The proposed hybrid system incorporates main components such as MHT, solar panel, batteries, converter, and electrical load profile. Based on these components plus electric cables, MHPS/solar PV/battery hybrid micro-grid system is formed. Therefore, the technical and economic investigation may be carried out. According to the power generation and electrical load profile, reliability and economic functions can be formulated. In other words, power generated from the system in the off-grid mode is required to reliably meet the anticipated electrical load profile at the lowest cost. This power generated from renewable sources is expressed as follows [49]:

$$P_{\text{ghybrid}} = P_{MHPS} + P_{\text{Solar}PV}, \tag{19}$$

where P_{ghybrid} is the total power generated from the hybrid micro-grid system (W), P_{MHT} is the power generated by one turbine (W), and $P_{\text{solar PV}}$ is the power output from solar PV (W). The power P_{ghybrid} produced by the renewable energy sources is required to generate sufficient power for the energy requirement plus losses. Energy storage, especially

batteries, is required to counteract the intermittent nature of renewable energy sources.

The power generation and overall performance of the entire hybrid system are an outcome of the operation of discrete component comprised in the system. Thus, the modelling of the components amalgamated in the system is as follows [54, 61].

7.1. Modelling of MHT. The potential of hydropower depends on the two primary input parameters such as water flow rate and water head. Generated power output from a single water turbine is computed using the following formula [40, 61]:

$$P_{MTH} = \frac{\eta_{MHT} \times \rho \times g \times H \times Q_{MHT}}{1000}, \quad (20)$$

where P_{MHT} is the power produced by a single turbine (kW), η_{MHT} is the water turbine efficiency (%), ρ is the density of water (kg/m^3), H is the water head (m), and Q_{MHT} is the water flow rate flowing into the water turbine (m^3/s).

As a matter of fact, hydropower turbine has practical and flow limits. It can operate effectively within certain flow ranges. If only one turbine is used in a hydropower system, three different operation stages can be distinguished as follows [40, 61]:

$$Q(t) = \begin{cases} 0, & if\ Q(t) \le Q_{Mn} \\ \text{Q}, & \text{if } Q_{Mn} \le Q(t) \le Q_{Mx}, \\ \text{Qmx}, & \text{if } Q_{Mx} \le Q\ (t) \end{cases} \quad (21)$$

where Q_{Mn} is the minimum flow rate, cutoff, below which electricity generation stops, and Q_{Mx} is the maximum flow rate a turbine can generate electricity, and Q (t) is the existing water flow rate.

The total power produced by the MHPS is calculated as follows [40]:

$$P_{MHPS} = N_{MHT} \times P_{MHT}, \quad (22)$$

where P_{MHPS} is the total power generated by the micro-hydropower system (ROR hydropower generation systems without pondage), N_{MHT} is the total number of water turbines, and P_{MHT} is the approximated power by one hydro-water turbine. Table 12 shows the summarized technical and fiscal specifications for MHT.

7.2. Modelling of Solar Photovoltaic (Solar PV). The solar PV system is designed by computing the suitable amount of solar panels. These solar panels transform solar radiation into direct current electricity. The solar panel is an arrangement of PV cells. Generated output per panel P_{SPV} at a given time in hours is defined as follows [54]:

$$P_{SPV}(t) = SOI(t) \times A_{SPV} \times \eta_{SPV}, \quad (23)$$

where SOI refers to the solar insolation in kW/m^2 at a given time t in hours, A_{SPV} is the surface area of a single solar panel (m^2), and η_{SPV} signifies the efficiency of solar panel (%). In this research work, solar panels are assumed to consist of the maximum power point tracking scheme and the effects of ambient temperature are negligible.

The total power output generated by a solar panel is estimated as follows [54]:

$$P_{\text{solar PV}}(t) = N_{SPV} \times P_{SPV}, \quad (24)$$

where $P_{\text{solar PV}}$ is the total power generated by solar panels in the proposed hybrid system. In this research work, the CSUN 410–144M Model of solar panel with peak power capacity of 410 W has been used in the performance analysis of the hybrid system. Table 13 illustrates the summarized technical and fiscal specifications for adopted solar PV.

7.3. Modelling of the Battery. Deep cycle batteries are mostly favoured for the storage of electricity for off-grid applications, particularly lead-acid batteries [54]. A detailed analysis of charging and discharging conditions of batteries is vital for the determination of finest capacity. The SOC (t) of a set of batteries at a time of one hour is related to SOC (t−1), the preceding condition of the charge, and the energy exhaustion or accrual state from t-1 to t [54].

If $P_{\text{ghybrid}}(t) = P_{solar\ pV}(t) + P_{MHT}(t) \ge PL(t)/\eta_{INV}$, generated electricity from given renewable energy sources is sufficient for meeting load demand P_{L} (t) plus extra energy kept in the batteries. At this state, the batteries are being charged. The SOC (t) is defined in the following expression [54]:

$$SOC(t) = SOC\ (t-1) \times (1-\sigma) + \frac{Pghybrid(t) - PL(t)/\eta INV \times \eta bc}{1000 \times Nb \times Cb}, \quad (25)$$

where η INV refers to the inverter efficiency, η_{cb} is the charging efficiency of the batteries, C_{b} is the nominal size of the battery in kWh, N_{b} is the quantity of batteries, and σ is self-discharging rate per hour.

Similarly, in case $P_{\text{ghybrid}}(t) < (\text{PL}(t))/(\eta_{\text{INV}})$, generated electricity from renewable energy sources is not sufficient and therefore does not satisfy electrical load demand. Energy stored in the batteries is discharged to balance the power supply and demand. For this condition, SOC (t) is expressed as follows [54]:

$$SOC(t) = SOC(t-1) \times (1-\sigma)\ \frac{PL\ (t)/\eta INV - Pghybrid\ (t)/\eta_{db}}{1000 \times Nb\ \times Cb}, \quad (26)$$

Table 12: Summarized technical and fiscal specifications for MHT.

Parameter	Specifics
Type of electric generator	Alternating current (AC)
Generated power by MHT	8–40 kW
Efficiency of the turbine	75%
Estimated head of turbine	13 m
Flow rate channel	160 l/s
Minimum flow rate	60%
Maximum flow rate	115%
Life span of the project	75 years
Charge for capital/unit	50, 000 USD
Charge for replacement/unit	43, 000 USD
Charge for operation and maintenance (O&M)/unit	5, 000 USD/year

Table 13: Summarized technical and fiscal specifications for adopted solar PV.

Parameter	Specifics
Company	CSUN
Panel model	CSUN 410–144M
Solar technology	Monocrystalline semi
Power capacity pmax	410 W′
Temperature range	−40°C–80°C
Efficiency	20.31%
Dimensions	1996 ∗ 991 ∗ 35 mm
Estimated area of the panel	1.98 m^2
Weight	22.9 Kg
Tc of voltage	−0.30%/° C
Tc of current	+0.06%/° C
Tc of power	−0.39%/° C
Life span of the project	25 years
Charge for capital/kW	2500 USD
Charge for replacement/kW	2500 USD
Charge for operation and maintenance (O&M)/kW	10 USD/year

Tc = temperature coefficient.

where η_{db} represents the discharging efficiency of batteries. Excessive charge ejection needs to be evaded in order to ensure a longer lifespan of batteries and SOC (t) should be subjected to the following restrictions [54]:

$$(1 - Max.DOD) \leq SOC\ (t) \leq 1, \tag{27}$$

where Max. DOD is the maximum depth of discharge. Alternatively, [40]

$$S0C_{Min} \leq SOC\ (t) \leq SOC_{Max}, \tag{28}$$

where SOC_{min} and SOC_{max} are lowest and highest value states of charge of the battery storage.

In this study, the Just Gel 12–100 Model of the battery is chosen for the performance analysis of hybrid system. Table 14 portrays technical and fiscal specifications for battery energy storage bank.

7.4. Modelling of the Converter. The converter is an electric device used for maintaining the flow of electricity in a renewable energy-based hybrid system. The converter comprises inverter and rectifier for conversion of DC power to AC and vice versa, respectively. Power from the output of converter should be higher than the maximum load. In this study, the power output of MHPS is AC that is connected to AC bus and power output of solar PV and battery energy storage bank is DC. Therefore, the inverter is essential for meeting electrical load demand. The rated power for the device is defined as follows [54]:

$$PL = P_{\text{inverter}} = \frac{P\text{max}AC}{\eta_{inverter}}, \tag{29}$$

where P_{inverter} is the rated power of inverter, $P\text{max}AC$ is the maximum AC electrical load, and η inverter is the efficiency of the inverter. Table 15 portrays the technical and economic parameters for adopted converter.

8. Optimization Models of the Proposed System Using GWO Algorithm

The MHPS/solar PV/battery bank hybrid system is optimized by taking into account different parameters such as loss of power supply probability (LEPP) for the system

Table 14: Technical and fiscal specifications for battery energy storage bank.

Parameter	Specifics
Company	Zhejiang Just
Type of the model	Just Gel 12–100
Weight	30.5 Kg
Operating temperature	25°C ± 3°C
Nominal voltage	12 V
Nominal capacity	100 AH
Peak capacity	100 AH
Efficiency of charging	90% (assumed)
Efficiency of discharging	95% (assumed)
Highest charging current	20 A
Highest discharging current	800 A
Depth of discharge (DOD)	80% (assumed)
Rate of self-discharge per hour	3%
Charge for capital	USD 200
Charge for replacement	USD 200
Charge for operation and maintenance (O&M)	USD 10
Life span of the project	15 years

Table 15: Technical and fiscal specifications for converter.

Parameter	Specifics
Rated capacity	40 KW
Inverter efficiency	90%
Inverter efficiency	85%
Charge for capital/unit	2000 USD
Charge for replacement/unit	2000 USD
Charge for operation and maintenance (O&M)/unit	0 USD/year
Life span of the project	25 years

reliability, total NPC, and COE. The sizing optimization of the system is obtained by the minimization of the NPC and also COE using the GWO algorithm. The optimization problem involves the formulation of objective function and constraints. In this study, the optimization of the objective function is performed under the restrictions such as upper and lower boundaries based on the amount of specified components and their power capacities, battery storage restrictions, and SPS for the technical and economic analysis of the system. The objective function under subjected limitations is formulated as follows [40, 49, 54].

8.1. Objective Function with Defined Constraints. The main purpose of the optimization is to attain the bottommost cost of hybrid power-generating system using the GWO method. The objective function is formulated to minimize the total NPC as follows [40, 49, 54, 62]:

$$\text{Total } NPC = \sum_{i=\left(N_{SoPV}+P_{MHPS+N_{BTR}}+N_{CONV}\right)} \frac{TANC_i}{CRP}. \tag{30}$$

The minimization of the total NPC according to the objective function is executed under the following constrictions as follows:

$$\begin{aligned} N_{SOPV}^{mini} \le N_{SOPV} \le N_{SoPV}^{maxi}, \\ P_{MHT}^{mini} \le P_{MHT} \le P_{MHT}^{maxi}, \\ N_{BTR}^{mini} \le N_{BTR} \le N_{BTR}^{maxi}, \\ N_{CONV}^{mini} \le N_{CONV} \le N_{CONV}^{maxi}, \\ LEPP^{mini} \le LEPP \le LEPP^{maxi}, \end{aligned} \tag{31}$$

where N_{SOPV}, N_{BTR}, and N_{CONV} are quantities of solar panels, batteries, and converters, respectively P_{MHT} is the power produced by micro-hydropower units forming P_{MHPS}. P_{MHPS} is power generated by MHPS from the tributary under study. Superscript maxi and subscript mini indicate maximum and minimum values of given parameters.

8.2. The Highest and Lowest Power Capacities of MHT. The water resource of Ikata tributary has been comprehensively discussed in Section 3. The model for MHPS has been executed in the HOMER Pro software. It should also be remembered that the range of the overall flow rate of the Ikata tributary starts from 160 l/s to 2.943 m^3/s. The minimum flow rate results during the dry season; thus, no power is produced from the tributary. So long, there is sufficient water flow rate during the rainy season and head, and the designed MHPS is capable of meeting the total actual load of 40 kW (greater than daily peak load). In this study, the turbine with an inconstant speed has been taken into consideration to maximize the use of the micro-hydropower potential within the given entire flow rate range and water head. The power capacity of the MHT can be estimated according to equation (23). Input parameters for sizing the MHT are tabulated in summary Table 12. In this speed adjustability of MHT, the minimum power capacity (P_{MHT}^{mini}) at time in hours is anticipated to be twenty percent of the highest turbine generation capacity (P_{MHT}^{maxi}), which balances the estimated actual load. The range of variable power from the MHT is expressed as follows [40, 54]:

$$8\,kW \le P_{MHT}(t) \le 40\,kW. \tag{32}$$

8.3. The Highest and Lowest Number of Components. The number of the solar photovoltaic panels, batteries, and converters may change to fulfil power supply requirement. The highest and lowest boundaries of the aforesaid components are presented in equations as follows [40, 63].

The maximum number of solar panels is computed as follows: in the first place, data for solar irradiance (1 kW/m^2), area (m^2), and efficiency of the model of the solar panel (20.31%) are tabularized in Table 13 to estimate power output produced by a single panel using equation (41). Therefore, [63]

$$P_{SPV} = 0.402\,kW. \tag{33}$$

The maximum number of panels for the generation power capacity balancing the expected actual load is approximated as follows [63]:

$$N_{SOPV}^{maxi} = \frac{PL}{P_{SPV}} = \frac{40\,kW}{0.402} \approx 100. \tag{34}$$

The maximum number of batteries can be estimated using the following equation. In the very beginning battery capacity, C_{BTR} is determined [63, 64] as follows:

$$\begin{aligned} c_{BTR} &= \frac{\text{Daily load } (Wh) \times D_D}{\text{System voltage } V \times \eta_{BTR} \times \max . DOD} \\ &= \frac{511100 \times 7}{48\,V} AH \\ &= \frac{74,\ 527\ AH}{48\,V} \\ &= 10,\ 647\ AH. \end{aligned} \tag{35}$$

The autonomy (D_D) of 7 days gives the total battery capacity equals 74, 529 AH. Max. DOD and overall battery efficiency are both assumed to be 100% So, the highest number of batteries can be computed using the following equation [63, 64]:

$$N_{BTR}^{maxi} = \frac{\text{Total esstimated capacity}}{\text{Nominal battery capacity}} = \frac{74,529\ AH}{100\ AH} \approx 746. \tag{36}$$

The protection against the excessive charge and over-discharge of the battery should be done for securing it from damage. The highest and lowest bounds of system storage may also be considered as boundaries. The boundaries are also directly proportional to the available number of batteries and SOC. The limits for battery bank capacities are given as follows [62]:

$$e_{BMI} \le e_B \le e_{BMA}, \tag{37}$$

where e_{BMI} and e_{BMA} are the least and greatest magnitudes of capacities of battery bank. These storage capacities are directly proportional to the available number of batteries and SOC and are computed in the following equations [62]:

$$e_{BMI} = \frac{N_{BTR} \times V_{N \times C_{BTR}}}{1000} \times SOC_{\min i}, \tag{38}$$

$$e_{BMA} = \frac{N_{BTR} \times V_{N \times} C_{BTR}}{1000} \times soc_{\max i}, \tag{39}$$

where V_N is the nominal voltage of the battery (V), C_{BTR} is the battery size (AH), and estimated values for $soc_{\min i}$ and $SOC_{\max i}$ are 0.2 and 1, respectively.

The highest number of converters is the calculated assumption that a single converter has rated capacity of about 2000 kW.

Thus, the highest number of converters is estimated as follows:

$$N_{CONV}^{\max i} = \frac{\text{Total actual load}}{\text{Power capacity of one converter}} = \frac{40\,kW}{2000} = 20. \tag{40}$$

8.4. The Highest and Lowest of LEPP. LEPP can be computed using (34) after the determination of the shortage of power supply using equation (35). For the straightforwardness of calculating the LEPP, in this study the equation (4) can be simplified into the following equation [49, 65]:

$$SPS = P_L - \left(P_{\text{ghybrid}} + P_{\text{battery bank}}\right). \tag{41}$$

If the power generated from renewable energy sources plus stored power in the battery bank satisfies all energy required for 100%, (4) rewritten as (42). LEPP can be estimated using equation (3) as follows [40, 49, 54, 65]:

$$SPS = P_L - 100\%\, P_L. \tag{42}$$

$$SPS = 0. \tag{43}$$

LEPP is determined by substituting (43) into (3), which equals zero. It implies that the system satisfies the power supply requirement when LEPP = 0.

Using the GWO method, the appraisal of the objective function according to the delineated constraints is made. Sizes of components are selected within the stated limits. The identification of viable patterns of the hybrid system is done in relation to the preset magnitudes of LEPP. The best possible solutions are repeated in the GWO method. The structure is updated to estimate the minimum total NPC.

The input parameters of PSO approach are shown in Table 11.

8.5. Methodology for Sizing the Proposed Hybrid Micro-Grid System Using the GWO Algorithm. Optimization parameters such as solar radiation, micro-hydro-energy source, and ambient temperature, electrical load data, and assumed constants (datasheet) were used as input variables. Decision variables such as A_{pv} and A_{MHT} were estimated by the GWO algorithm for individual iteration. Then, the power from solar PV (P_{pv}) and micro-hydro-power systems (P_{MHT}) was estimated with variable inputs, assumed constant inputs, and approximated decision variables. To meet the energy requirements, power is considered to be generated by MHT (MHPS) and solar PV. However, if generated electricity from aforesaid energy sources does not satisfy the load demand, stored energy from batteries is used to balance power supply and demand. The standards of LEPP were calculated for the total time in hours for a year equivalent to 8760 hours. In case any value of LEPP has been greater than the prescribed one, the decision values were then recalculated. If there was no deviation from the predetermined value, an optimization procedure was implemented. Then, the optimal size of the hybrid system is obtained corresponding to the lowermost total NPC (also COE), which is computed at LEPP equals zero when no shortage of power supply. Table 16 indicates miscellaneous specifications for the GWO technique.

Table 16: Miscellaneous specifications for the GWO technique.

Parameters	Value
Population size of grey wolves n	12
Maximum number of iterations, $\text{iter}_{\max}$	100
Penalty factor, PP	10^{10}
Dimension, d	2
Inflation rate	5%
Interest rate	3.04%
Life span of the project	25 years
Highest capacity of micro-hydro-turbines $P_{MHT}^{\max i}$	40 kW
Highest quantity of solar panels $N_{SoPV}^{\max i}$	100
Highest quantity of batteries $N_{BTR}^{\max i}$	764
Highest quantity of converters	20
Highest LEPP, $LEPP^{\max i}$	0.07

9. Results and Discussions for the Technical and Economic Analysis of Proposed Hybrid Renewable Energy-Based System

The optimal hybrid configuration indicates micro-hydro-energy, solar energy, energy storage system, and power conditioning devices. MHPS consists of main components such as water turbine and generator. SPVS comprises solar panels, charge controllers, and batteries and converter. HOMER has been used to optimize the hybrid energy system. Figures 12 and 13 indicate optimal configuration for the proposed off-grid hybrid energy-based system and schematic depiction for off-grid hybrid micro-hydro- and solar system, respectively.

The optimal hybrid energy system has been simulated using HOMER Pro software according to the technical and economic details. Simulation results are of two kinds, that is, technical and economic results. In this simulation load, the following has been used as a dispatch strategy. Therefore, the simulated results of the optimized renewable energy system are presented as follows.

In this section, the results of electrical power analysis that include solar and micro-hydro-resource outputs, electrical power output, electricity generation, and performance of individual systems/components within the proposed hybrid energy system have been presented. The system comprises MHPS, solar photovoltaic panels, batteries, and converter. The proposed hybrid system will specifically use Natel-free jet (MHT), generic solar photovoltaic, and PowerSafe SBS 1500 (batteries). The designed system has the capacity of supplying electricity to the proposed site of peak load of 30.3 kW for a period of 25 years and interest rate of 3.04%. The installed generation capacity of the designed system should be equal to or greater than the daily peak power in order to meet the load demand. To account for losses by ensuring reliability and availability of power, the designed system should have greater capacity than the daily peak power. Therefore, the installed generation capacity is approximated to be 48 kW. Technical specifications of the system architecture are given in Table 17. The proposed hybrid energy system generated excess electricity, unmet electrical load, and capacity shortage for the proposed hybrid energy system which are 128 483 kWh/year, 0 kWh/year, and 0 kWh/year, respectively. Table 18 indicates the generation, consumption, and renewable energy utilization. Miscellaneous technical and economic details for all components in the hybrid energy system are presented in Tables 19 and 20, respectively.

Additional related information or parameters for MHPS are that the lifespan for MHPS is assumed to be 75 years. The estimated theoretical generation power capacity of the MHPS is 375 kW considering the water head of 13 metres and flow rate of 2943 litres per second. The minimum flow rate is 160 litres per second. The maximum and minimum values of water flow of the tributary are depicted in Figure 4. Generally, the maximum power-generated output of MHPS occurs during the rainy season, particularly in January. However, practically, the power output drops down depending on the rainfall-dependent water flow profile of the tributary under study and other technological aspects. Technological factors are water discharge rates and the efficiency of the turbine. The maximum and minimum discharge rates are 115 and 60 percentages, respectively. The efficiency of MHT in this study is considered to be 75%. During the dry season, that is, in the absence or shortage of water flow in the tributary, MHPS does not produce any power output. In the period starting from June to September, no power is generated from the MHPS. Figures 14 and 15 show the annual daily profile of micro-hydropower output and annual monthly averages of hydropower output, respectively.

Specific technical and economic details are also presented in Tables 19 and 20. As indicated in Figure 6, the area of proposed site has sufficient solar energy potential. This potential alongside battery energy storage assures the availability and reliability of power supply. The lifespan of SPVS is approximated to be 25 years, and the system is ground-mounted with a derating factor of 90%. It is capable of providing power output throughout both rainy and dry seasons for the proposed hybrid energy systems. Figure 15 displays the annual daily profile of solar photovoltaic power output (Figure 16).

Similarly, the additional technical and fiscal details for battery energy system are also tabulated in Tables 19 and 20. The life span of battery has been assumed to be 15 years; the battery energy system employs a total of 288 batteries of PowerSafe SBS 1500. The nominal voltage for the battery is 12 volts, and bus voltage is 48 volts (there are 4 batteries in series/per string and 24 parallel strings). The battery bank has the autonomy of 163 hours. State of charge (SOC) is almost constant all the time except in the period dry season from June to September. Figure 17 shows the annual monthly average state of charge (SOC) of battery bank.

The technical and financial details for converter are given in Tables 17, 19, and 20. The maximum capacity for the converter is 40 kW. The lifespan for the inverter is assumed to be 25 years. The performance of inverter is assumed to be 90%. The efficiency of rectifier and inverter is 90% and 85%, respectively.

The whole hybrid energy system produces an annual peak power output equivalent to 40 kW except in the period of June and September in which the major source of electricity is solar radiation. Figures 18 and 19 indicate the annual daily profile of the total renewable energy output and

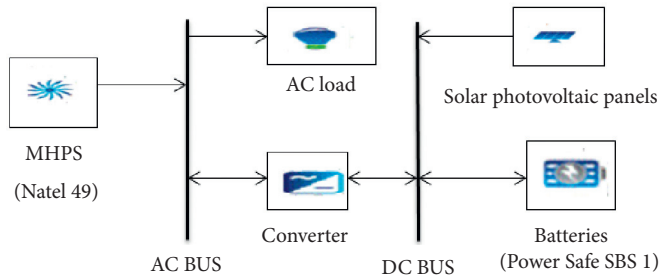


Figure 12: Optimal configuration for the proposed off-grid hybrid energy-based system.

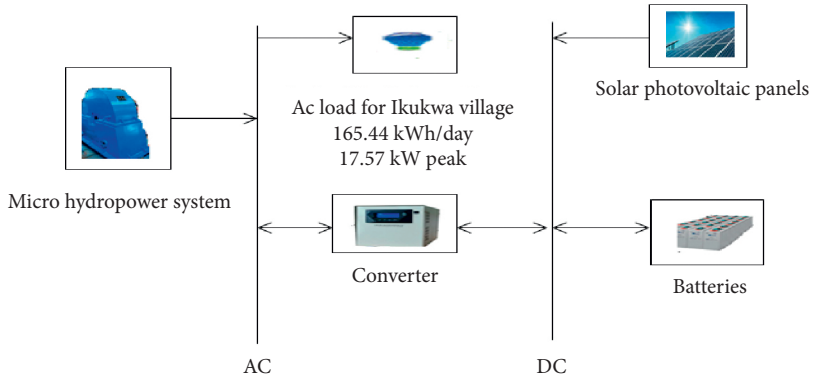


Figure 13: Schematic depiction for optimal configuration for the proposed off-grid hybrid energy-based system. Simulation results for technical analysis by HOMER software.

Table 17: System architecture.

Component	Capacity in kW/quantity in numbers
Generic flat-plate photovoltaic panels	40 kW
Micro-hydro-turbine/Natel-free jet	18.4 kW
Converter	40 kW
Batteries/PowerSafe SBS 1500	96 nos

Table 18: Generation, consumption, and renewable energy utilization.

Characteristics/factors	Unit/value	Percentage
Production	KWh/year	
Solar photovoltaic	75 804	39.8
Micro-hydro-turbine	114 431	60.2
Total	190 235	100
Consumption	KWh/year	
AC primary load	60 386	100
DC primary load	0	0
Deferrable load	0	0
Total	60 386	100
Quantity	KWh/year	
Excess electricity	128 483	67.5
Unmet electrical load	0	0
Capacity shortage	0	0
Quantity of renewable sources	Value	Unit
Renewable fraction	100	%
Max. renewable penetration	11 347	%

annual scaled data profiles, respectively. During the rainy season, MHPS has a higher generation capacity than SPVS. SPVS with battery energy storage gives power in all seasons of the year within the proposed hybrid system. In dry season, SPVS is the only alternative energy source of power. Similarly, in the period of dry season battery energy storage

Table 19: Economic input parameters for the optimized hybrid renewable energy system.

Component	Capacity/quantity (kW/Nos)	Capital cost ($)	Replacement cost$	O&M cost ($/year)
MHT/Natel 49	18.4 kW	50 000	43000	110
Solar panels	40 kW	2500	2500	10
Batteries	288 nos	2500	2500	120
Converter	40 kW	1000	1000	110

MHT = micro-hydro-turbine.

Table 20: Annualized costs by component.

Type of cost	Components of optimized hybrid renewable energy system				System
	MHT	Flat solar PV	Batteries	System converter	
Capital ($)	3867.72	193.39	1, 547	77.35	5685.55
Replacement ($)	0	0	0	32.82	32.82
O&*M* ($)	5000	10	960	110	6080
Fuel ($)	0	0	0	0	0
Salvage ($)	531.22	0	281.67	6.18	819.06
Total ($)	214	203.9	225.42	214	10979.4

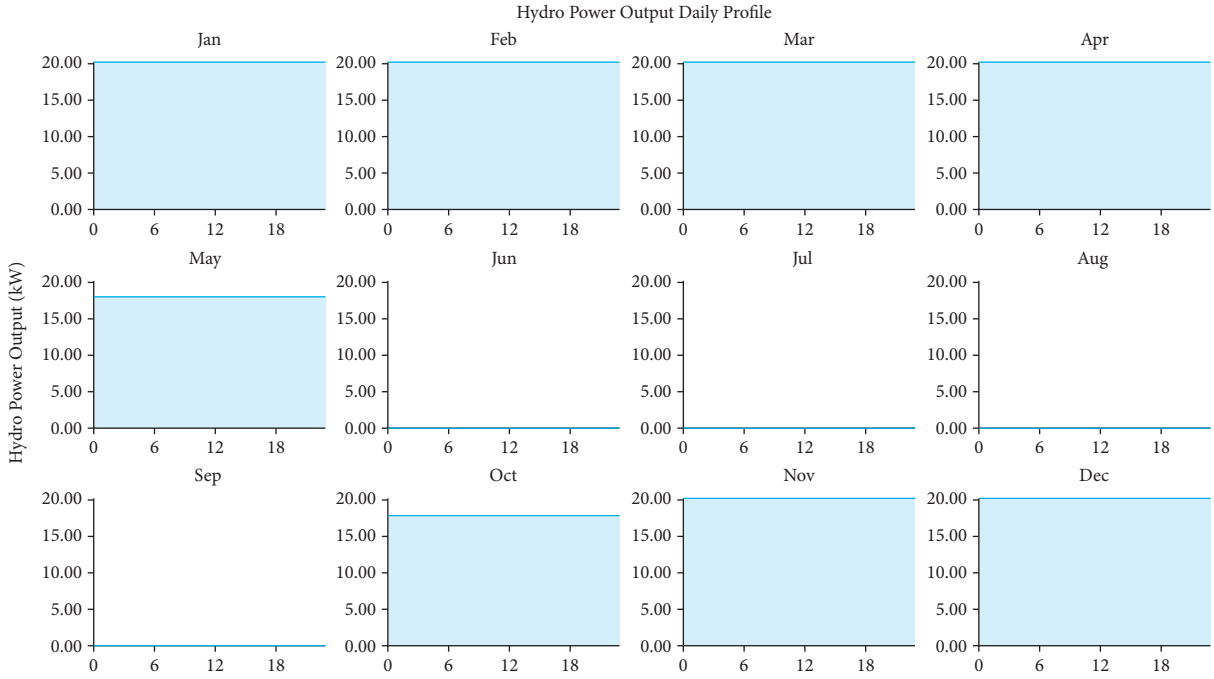


Figure 14: Annual daily profile of micro-hydropower output.

demonstrates the highest input power to the system. Figures 20 and 21 illustrate the annual monthly electricity production and hourly power production of the proposed system, respectively. The annualized cost by components of MHT has the highest cost, followed by the cost of batteries, and then, the cost of generic flat solar and the cost of converter are the least. Figure 22 and Table 19 both indicate annualized cost by component in the hybrid system.

This section of this article presents an economic analysis of a 100% renewable energy system consisting of solar photovoltaic panels, micro-hydro-turbine, batteries, and converter. This analysis has been carried out by considering the aforesaid period of the project. The types of annualized and net present costs by components have been analysed. Various types of costs such as initial cost, replacement cost, operation and maintenance cost, fuel cost, salvage cost, and inflation rate have been considered to be 3.04%. Table 19 shows economic input parameters for the optimized hybrid renewable energy system. The results for the total net present cost, operating cost, and cost of energy are $ 141, 397.10, $ 5293.76, and 0.1818/kWh, respectively. Annualized costs by components are indicated in Figure 22 and Table 20, respectively.

9.1. Limitations of the Proposed Hybrid Energy System. Apart from the merits of including the flow of the river system, solar PV and batteries in the proposed hybrid system

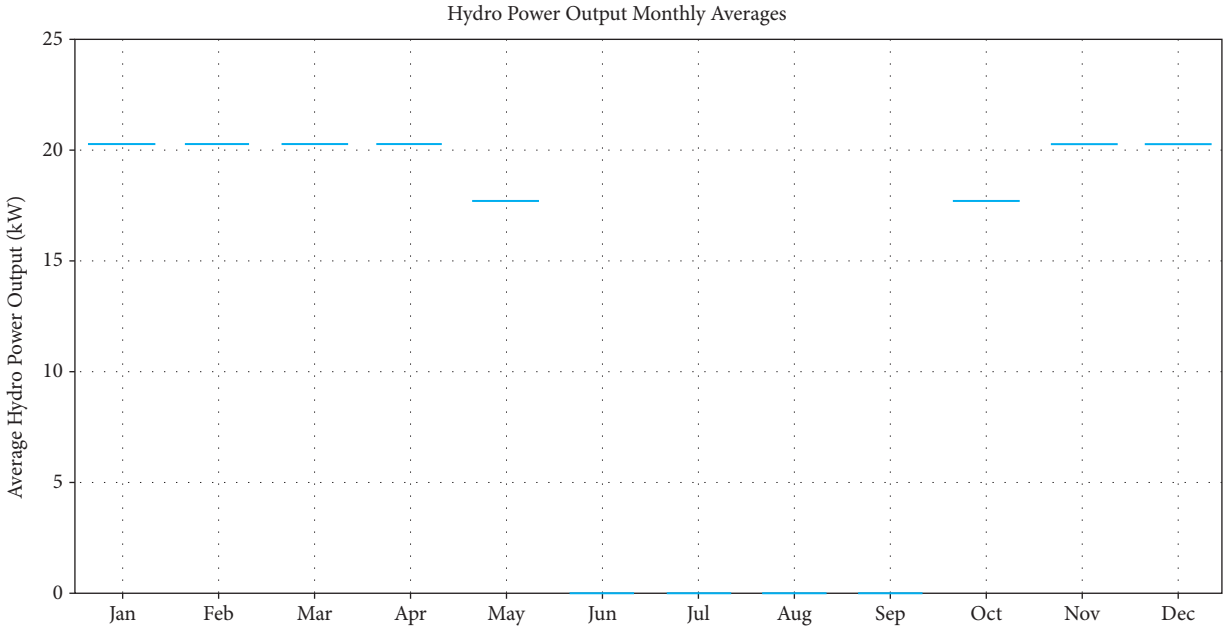


Figure 15: Annual monthly averages of hydropower output.



Figure 16: Annual daily profile of solar photovoltaic power output. Simulation results for economic analysis by HOMER software.

such as reduced investment cost, running cost (water source and solar energy are free and renewable), time of construction, negative impacts of environment, and power oscillations (inter-day and seasonal variability), the limitations of the hybrid system are described as follows [66]:

(i) The operation of the hydropower system in nature is a weather-dependent project. Better performance of the flow of the river system depends on the high precipitation in the area of study. The operation of the system may be drastically affected in the critical dry season. A suitable operation strategy is necessary to avoid abnormalities, particularly during the dry period. The analysis of this specific system does not include control strategy to address the challenge of uncertainty and unpredictability of water flow for the inter-daily, rainy, and dry periods [66].

(ii) The analysis of the system does not include the dynamics of the solar cells and hydropower excitation system [66].

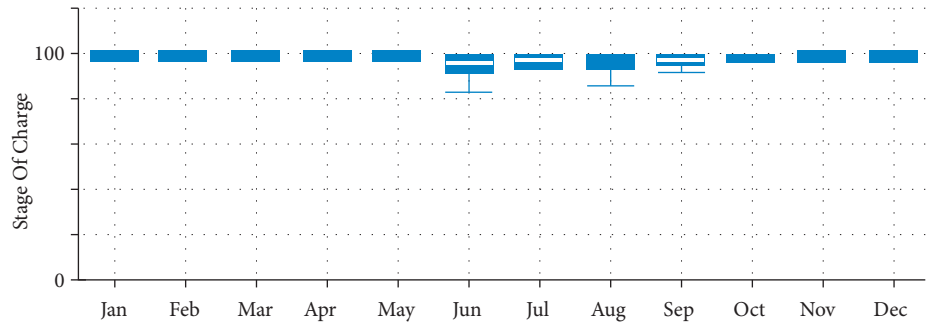


Figure 17: Annual monthly average SOC of battery bank.

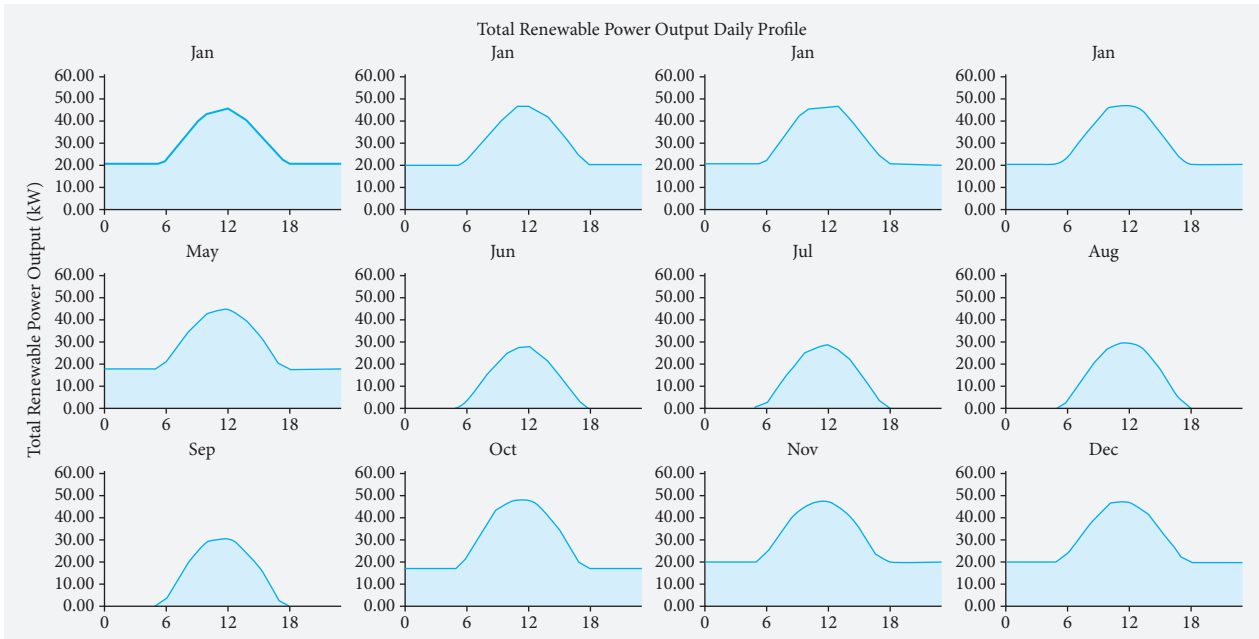


Figure 18: Annual daily profile of the total renewable energy.

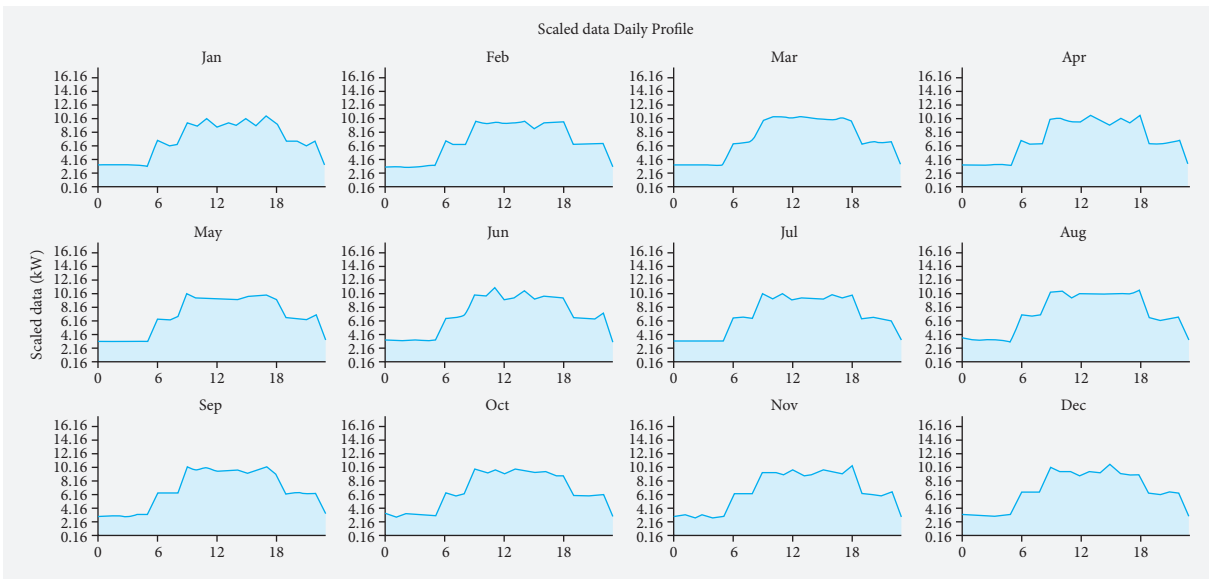


Figure 19: Annual scaled data load profile.

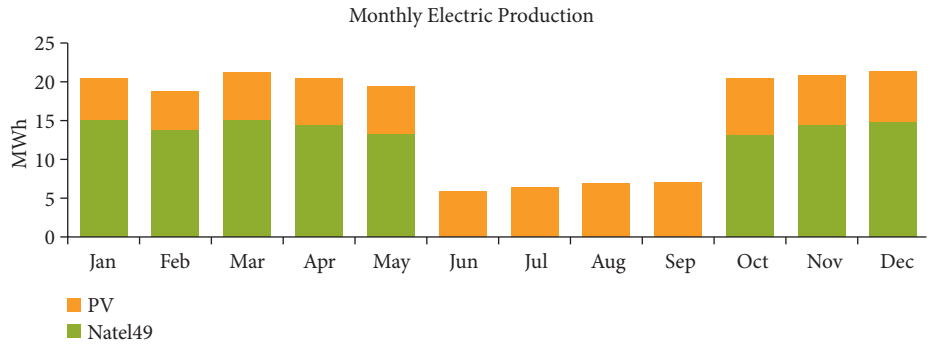


Figure 20: Annual monthly electricity production.

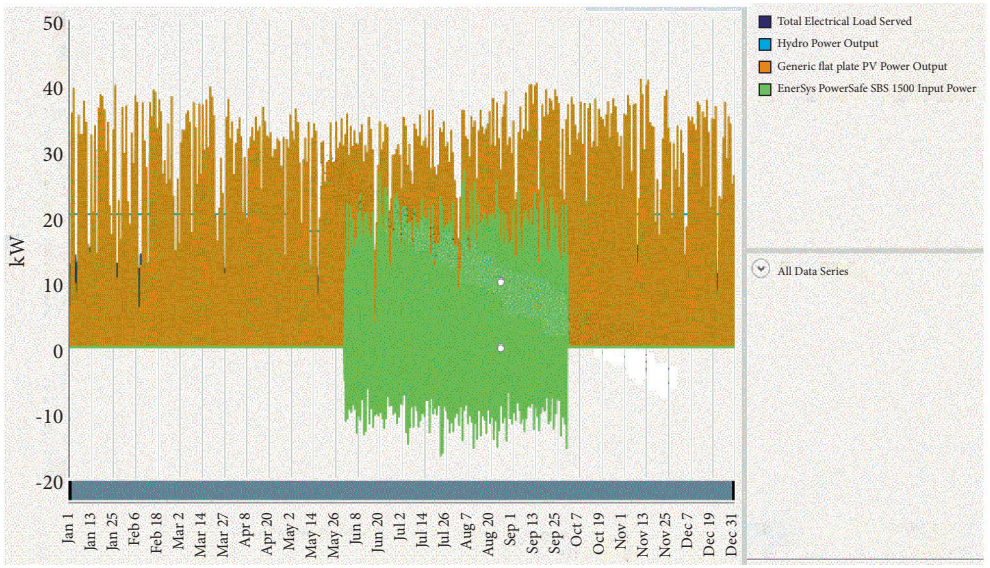


Figure 21: Hourly power production of the proposed system.

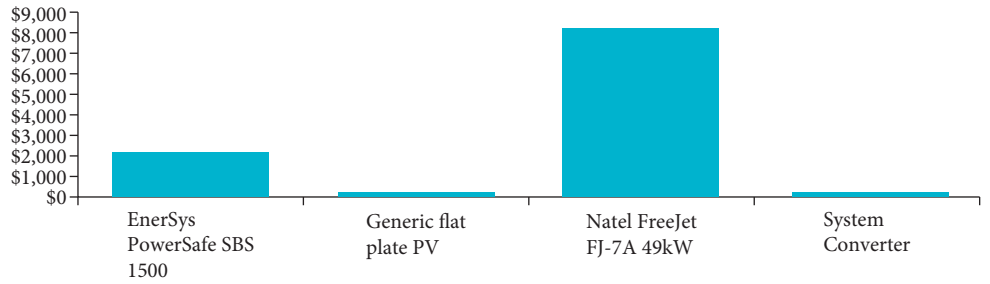


Figure 22: Annualized costs by component.

(iii) The dynamic response of hydro-energy system due to great inertia of water turbine owing to the requirement of the battery energy storage has not been considered in the optimization [56].

(iv) Fixed efficiency has been used in the optimization process of the system. As water injection from the water turbine affects the efficiency of run of over the river power plant, no control has been included to balance with variable electrical load. This is not practical as far as water flow is variable due to the fluctuating hydrological situations.

(v) Application of traditional methods for data measurements may provide inaccurate and unreliable data. In addition, data gathering does not include geological information [66].

(vi) The analysis of the proposed hybrid system is not grid-connected. The analysis with the addition of grid connection to the hybrid energy system would

be helpful for the minimization of cost incurred by increasing the number of batteries in off-grid decentralized systems [66].

(vii) Complicated control and limited lifecycle of the battery are subjected to the rainy and dry seasons [66].

(viii) There is no practical validation of the proposed system to gauge the performance of the system in the real environment due to the economic challenges and limited time.

(ix) The profile characteristics of the run-of-the river pattern may be affected by weather variations as a result of drastic effect of global warming [66].

(x) The analysis of the proposed energy system does not involve the environmental component of sustainability. The recommendation that has been described in the conclusion is only based on authors' observation in the area of study [67, 68].

For appropriate performance analysis of the hybrid system, the artificial intelligent-based control strategy should be implemented.

9.2. Assessment Criteria of the Proposed Hybrid Energy System. The proposed hybrid system consisting of MHT, solar PV, and batteries has been optimized using the monthly average flow of water from the tributary under study and global solar radiation. The tributary is located in one of the rural areas of Mbeya Region in Tanzania having prolonged high rainfall. Also, the tributary is well elevated lying in Mount Mbeya having sufficient water flow particularly during the wet season mainly contributed by high rainfall and a few months after the rainy season. The water flow of the tributary is discharged into the River Shongo, which provides water for the livestock and local irrigation schemes of agriculture at Ikukwa Village. The water flow from the tributary is deemed to have a potential of contributing energy to the proposed system in the wet season. The area under study is also exposed to the sufficient coverage of sunlight during the dry season. Solar energy can be harnessed to generate electricity in the hybrid system. However, sunlight is absent at night time and during the rainy season, and therefore, battery energy storage is necessary for covering the excess or shortfall of power supply and for ensuring the fast response of off-grid hydro-energy system. In other words, solar energy source is categorized by uncertainties and huge fluctuations, which can impact adversely the system steadiness and consistency requiring batteries for energy storage. In general, the hydro-energy source provides high power, reliable in wet season, and dries up in the dry season, while solar energy source is intermittent, and thus, these types of energy sources are complementary for ensuring the 100% availability of power supply to meet the load for the whole year. These complementary energy sources are sustainable, renewable, free, and environmentally friendly. The proposed hybrid energy has been analysed using collected average data of available renewable energy sources and electrical demand. Based on the authors' practical experience, literature review and decision-making methods via up-to-date advancement of using a computer application and AI optimization techniques of the potential of hybrid system have been evaluated, and assessment criteria are summarized as follows [66]:

(i) The location of study has a high potential of water flowing in the tributary during the wet season and sufficient solar energy source for electricity generation. The energy sources are sustainable and complementary renewable energy sources [67, 68].

(ii) For the purpose of reliability improvement of the system and cost minimization of batteries, hybridization of renewable energy sources is the best option. Solar and hydro-energy sources are natural-free sources and are also ecologically friendly, and thus, the hybrid power plant combining the ROR power system and solar PV is cost-effective and sustainable. The configuration of the plant has been defined by the lowest NPC [67, 68].

(iii) Generated power output of the run-of-over river scheme in the optimal hybrid system was estimated using water flow rate, water head, and fixed efficiency. In reality, the dynamics of the hydrological situation always varies over time. Therefore, one of the drawbacks of this method is that it overestimates the actual generated power, particularly run-of-the river systems with low water head due to the limitations of the turbine release, head, and efficiency deviations [67, 68].

(iv) Power capacities of solar PV and batteries have been estimated according to the monthly profiles of the streamflow, solar radiation, and electrical load profile. The tributary profile was obtained using traditional methods of measurements. Solar radiation data were directly obtained from NASA on the monthly basis, while batteries have been sized according to hourly characteristics for minimization of the hourly fluctuations of solar PV in day time [67, 68].

(v) Energy requirement was estimated based on rated power capacities of anticipated appliances, and load profile was computed according to the socioeconomic life and norms of the area of study. In the optimization processes of the proposed system, it was assumed that generation capacity greater than the peak maximum accounts for the power losses [67, 68].

9.3. Optimization Results of the Proposed System Using AI Algorithms. As it has been mentioned before, the average daily electricity consumption is 511.1 kWh/day (baseline), peak power capacity is 30.31 kW, and load factor is 0.71. For the hybrid electric system to supply sufficient power to satisfy the load, generated electricity should be equal to or greater than the highest power capacity. The proposed system has been optimized when LEPP equals zero denoting no deficiency of power assuming that the hybrid energy system supplies power at 100%. Optimization results that are processed by AI techniques such as PSO, GWO, and GWO-

Table 21: Comparative analysis of optimization results of the soft computing and AI methods for the optimized MHPS/solar PV/battery system (LEPP = 0).

Financial parameters	HOMER	PSO	GWO	GWO-PSOHD
NPC ($)	141 397.76	95 167.21	92 472.82	91,854.10
COE ($/kWh)	0.1818	0.1185	0.1182	0.1181

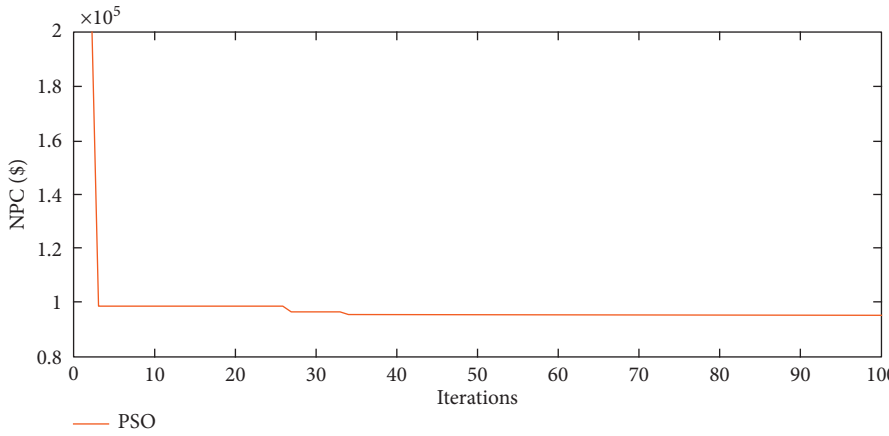


Figure 23: NPC optimization for the MHPS/solar PV/battery system using the PSO algorithm.

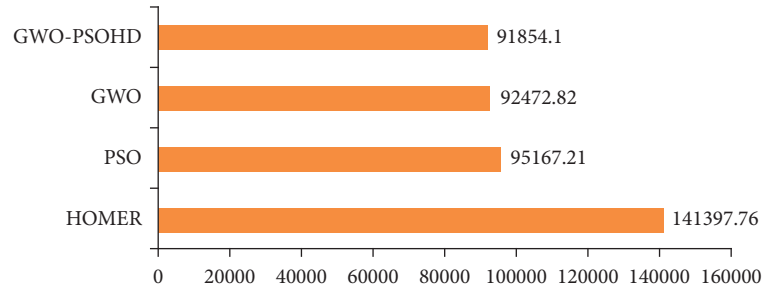


Figure 24: NPC for the MHPS/solar PV/battery system by soft computing and AI methods.

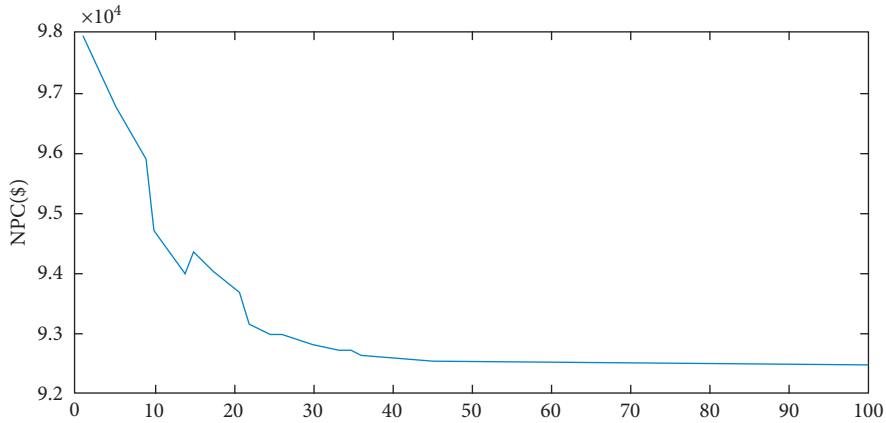


Figure 25: NPC optimization for the MHPS/solar PV/battery system using the GWO algorithm.

PSOHD algorithms show the reduction in total NPC and corresponding levelized COE in reference to the soft computing method (HOMER application). Table 21 shows the comparison of the optimization results of the soft computing and AI methods for the optimized MHPS/solar PV/battery system (LEPP = 0). The AI techniques are implemented in MATLAB, and optimal results are presented as follows.

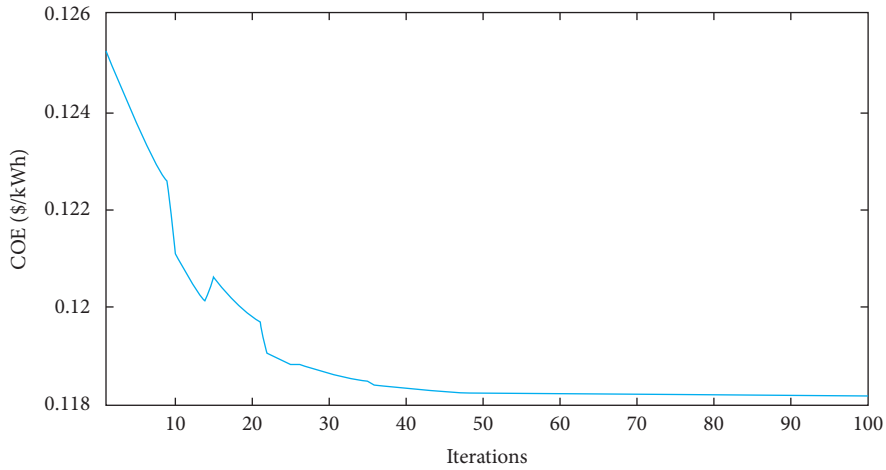


Figure 26: COE optimization for the MHPS/solar PV/battery system using the GWO algorithm.

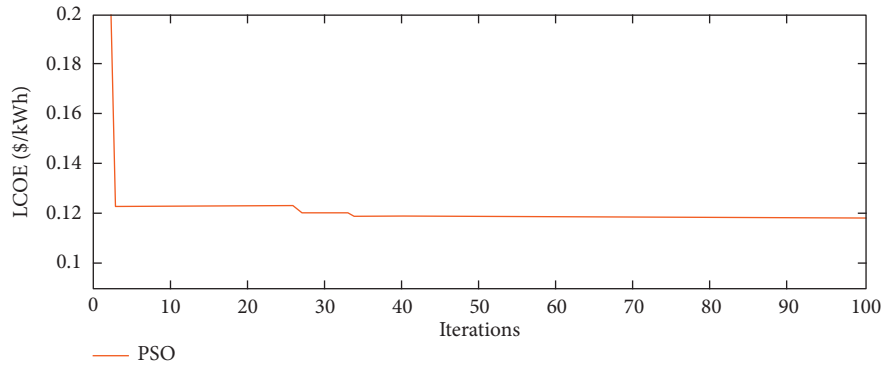


Figure 27: COE optimization for the MHPS/solar PV/battery system using the PSO algorithm.

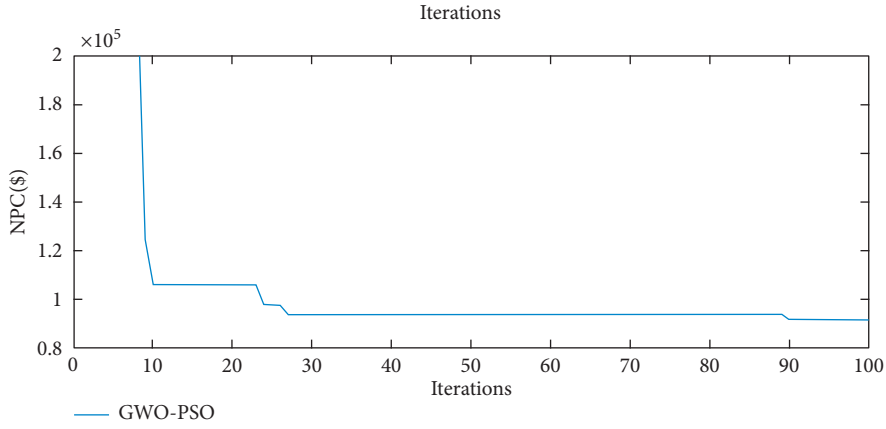


Figure 28: NPC optimization for the MHPS/solar PV/battery system using the GWO-PSOHD algorithm.

9.4. Results of PSO Implementation. Among the above named three AI optimization techniques, the PSO method has indicated the greatest NPC after HOMER software. The PSO implementation gives accumulated NPC and its corresponding levelized COE, which are $ 92,472.82 and $ 0.11 82/kWh, respectively. These categories of financial standards are lower in the PSO method than in the HOMER platform. The NPC and COE are more minimized nearly 67.3%. It denotes that the PSO method offers the superior solution in terms of reduction in cost in comparison with HOMER software although it has the greatest cost in comparison with those costs achieved by the GWO and GWO-PSOHD optimization techniques. Figure 23 describes the NPC optimization results for the MHPS/solar PV/battery system

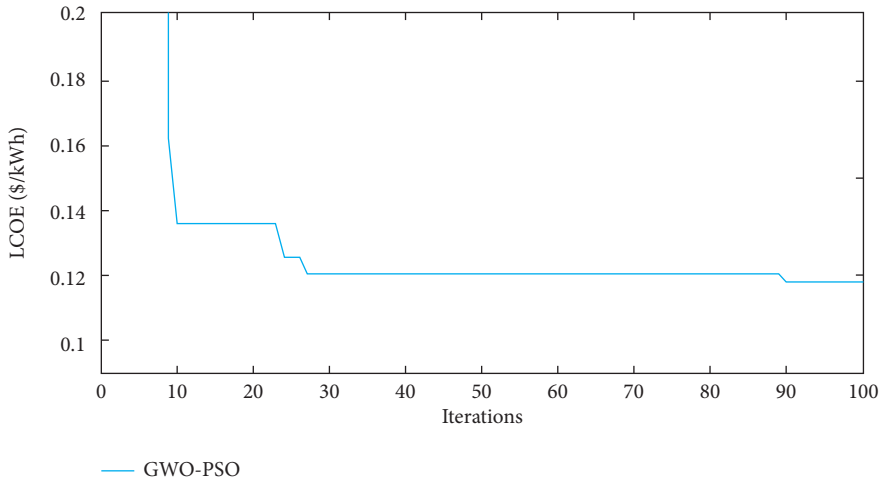


Figure 29: COE optimization for the MHPS/solar PV/battery system using the GWO-PSOHD algorithm.

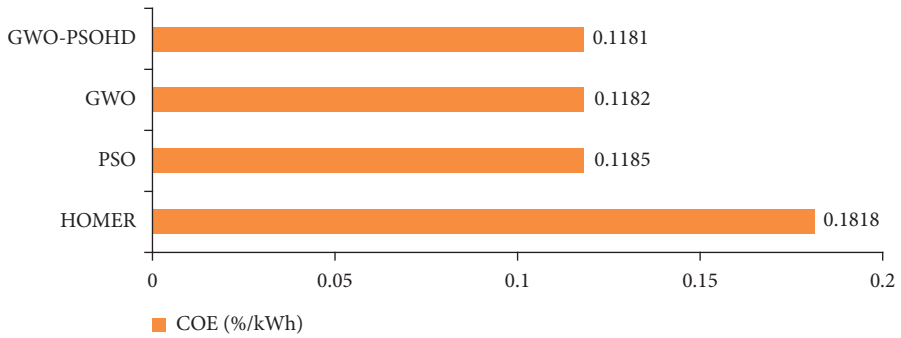


Figure 30: COE for the MHPS/solar PV/battery system by soft computing and AI methods.

using the PSO algorithm. Furthermore, Figure 24 illustrates COE optimization for the MHPS/solar PV/battery system using the PSO algorithm.

9.5. Results of GWO Implementation. GWO implementation minimizes further accrued NPC and levelized COE of the proposed off-grid hybrid renewable energy-based system having the magnitudes, which are less than that achieved from the PSO method. The optimally designed system using GWO provides NPC and COE, which are $ 92,472.82 and $ 0.11 82/kWh, respectively. These types of economic values are less in the GWO technique than in the HOMER. Both NPC and COE are further reduced up to around 65%. The GWO techniques offer an intermediate cost between the PSO and GWO-PSOHD techniques having reduced cost in comparison with HOMER software. Figure 25 depicts the NPC optimization results for the MHPS/solar PV/battery system using the GWO algorithm. Furthermore, Figure 26 illustrates COE optimization for the MHPS/solar PV/battery system using the GWO algorithm (Figure 27).

9.6. Results of GWO-PSOHD Implementation. The research of this study has employed the GWO-PSOHD algorithm to optimize the proposed energy system. Similarly, this technique has been used to solve the formulated objective function by the minimization of NPC. The number of iterations and population size is 100 and 50, respectively. Optimal results have indicated that the hybrid algorithm provides the least value of NPC in comparison with other aforementioned optimization methods. Both NPC and COE are more reduced up to about 64.96%. Figure 28 represents the NPC optimization results for the MHPS/solar PV/battery system using the GWO-PSOHD algorithm. Furthermore, Figure 29 illustrates COE optimization for the MHPS/solar PV/battery system using the GWO-PSOHD algorithm.

Moreover, Figures 24 and 30 portray the total NPC and COE for the MHPS/solar PV/battery system of the soft computing and AI optimization methods, respectively. Table 21 illustrates the comparative analysis of the optimization results of the soft computing and AI methods for the optimized MHPS/solar PV/battery system (LEPP = 0).

10. Conclusions, Recommendation for the Long-Term Sustainability of the Tributary under Study, and Future Research Work

The techno-economic analysis of unexploited potential of a nonperennial tributary incorporated with solar energy source and batteries has been carried out. A case study has been made

at Ikata tributary located at Ikukwa Village, Mbeya Rural District, Mbeya Region, Tanzania. The main objective is to examine the potential of using the unused tributary hybridized with solar photovoltaic and batteries to meet electrical load demand for small communities without connection to the grid power network. Data for the estimation of water flow rate and head for Ikata tributary have been achieved by direct measurement using traditional methods.

In the micro-hydro/solar photovoltaic/battery system, MHPS immensely indicates a high positive impact in terms of power output. The overall proposed hybrid energy system provides sufficient power, which is mainly contributed by MHPS during the extended period of rainy/winter season. Power capacity for MHPS in the hybrid energy system drops down to zero due to the limitation or absence of water flow in dry period. Within this proposed hybrid energy system, SPVS with batteries provides electricity during rainy and dry seasons. However, SPVS with batteries enormously offers electricity during dry/summer seasons with limited or absence of power output from MHPS.

Three major implications can be made about this research work. First, the hydro-energy source from the tributary has a high abandoned potential of energy, especially in the rural expanses with high and prolonged rainfall. Second, this proposed hybrid energy system is economically feasible with the lowest total NPC and COE. Third, the optimal design of the proposed hybrid system produced by AI optimization techniques (PSO, GWO, GWO-PSOHD) is more cost-effective than that executed by soft computing technique (HOMER Pro software). In the group of AI optimization approaches, the PSO technique has the highest price, while the cost acquired by the GWO-PSOHD technique is the lowest, and the cost acquired by the GWO method lies between the two.

In the survey made by the authors at the site of the project, much deforestation was observed. Villagers get firewood for cooking food and heating purposes by cutting trees in Mbeya Mountains. The ongoing deforestation should be stopped for achieving the long-term sustainability of Ikata tributary and other hydro-energy sources [23, 69]. The action of stopping the deforestation can also sustain the performance of MHPS in future.

In this study, the traditional approach in direct measurements of power has been applied and used for the mathematical analysis of theoretical power of MHPS. Some of the parameters obtained by such approach, for instance, water flow rate, have been used as a reference during the simulation process by HOMER Pro software. Therefore, in future the research can be furthered by improved methodological approaches including more modern AI methods while integrating other energy sources with the system under investigation. The detailed techno-economic feasibility analysis can be implemented using advanced methods of measuring techniques, geological survey, and grid integration to support the national power network from TANESCO. Also, this study can be expanded by implementing sensitivity analysis under the varied LEPP on the assumption that the hybrid renewable energy-based system does not produce power at 100% [70].

Data Availability

Data are available on the request from the corresponding author.

Conflicts of Interest

The authors declare that there are no conflicts of interest.

Authors' Contributions

IJM collected the data in Tanzania, performed the analysis, and wrote the manuscript. MR and NK provided valuable inputs for better output of research work. All authors read and approved the manuscript.

Acknowledgments

This research was part of the PhD studies supported by the Indian Council for Cultural Relations (ICCR) under the Government of India (F. No. ASS/51/2018). The authors are indebted to the management of Delhi Technological University for the support. Appreciation is also extended to the Regional Commissioner's Office in Mbeya Region for the permission of allowing authors to carry out the research work in Mbeya Rural District. The authors are also delighted to thank the administration office of Ikukwa ward and the residents for their support during the site survey and data collection.

References

[1] T. Zhang, X. Shi, D. Zhang, and J. Xiao, "Socio-economic development and electricity access in developing economies: a long-run model averaging approach," *Energy Policy*, vol. 132, pp. 223–231, 2019.

[2] T. Jacob, "Competing energy narratives in Tanzania: towards the political economy of coal," *African Affairs*, vol. 116, no. 463, pp. 341–353, 2017.

[3] O. J. Mdee, T. K. Nielsen, C. Z. Kimambo, and J. Kihedu, "Assessment of hydropower resources in Tanzania," *A review article*, 2018.

[4] V. Sridharan, E. P. Ramos, C. Taliotis, and M. P. I. V. Howells, "Vulnerability of Uganda's electricity sector to climate change: an integrated systems analysis," *Handbook of Climate Change Resilience*, vol. 10, no. 1, pp. 1–30, 2019.

[5] O. K. Bishoge, G. G. Kombe, and B. Mvile, "Community participation in the renewable energy sector in Tanzania," *International Journal of Sustainable Energy Planning and Management*, vol. 28, pp. 121–134, 2020.

[6] A. Kalair, N. Abas, M. S. Saleem, A. R. Kalair, and N. Khan, "Role of energy storage systems in energy transition from fossil fuels to renewables," *Energy Storage*, vol. 3, no. 1, p. 135, 2020.

[7] E. Muh and F. Tabet, "Comparative analysis of hybrid renewable energy systems for off-grid applications in Southern Cameroons," *Renewable Energy*, vol. 135, pp. 41–54, 2019.

[8] S. Zakaria, "The political economy of decentralized approaches to rural electrification in Tanzania: implications for agro-industry co-generation projects," *IIIEE Masters Theses*, 2016.

[9] A. Chauhan and R. P. Saini, "Discrete harmony search based size optimization of Integrated Renewable Energy System for remote rural areas of Uttarakhand state in India," *Renewable Energy*, vol. 94, pp. 587–604, 2016.

[10] O. Bishoge, L. Zhang, and W. Mushi, "The potential renewable energy for sustainable development in Tanzania: a review," *Cleanroom Technology*, vol. 1, no. 1, pp. 70–88, 2018.

[11] U. F. Melkior, J. Tlustý, and Z. Müller, "Use of renewable energy for electrification of rural community to stop migration of youth from rural area to urban: a case study of Tanzania," *Dependability Engineering*, 2018.

[12] F. A. Anyogu, C. A. Eme, and J. A. Ogbodo, University-Led Knowledge and Innovation for Sustainable Development, 2021.

[13] M. Costa, A. Buono, C. Caputo et al., "The "INNOVARE" project: innovative plants for distributed poly-generation by residual biomass," *Energies*, vol. 13, no. 15, p. 4020, 2020.

[14] I. Haq, K. Qaisar, A. Nawaz et al., "Advances in valorization of lignocellulosic biomass towards energy generation," *Catalysts*, vol. 11, no. 3, p. 309, 2021.

[15] Z. Aslam, H. Li, J. Hammerton, G. Andrews, A. Ross, and J. C. Lovett, "Increasing access to electricity: an assessment of the energy and power generation potential from biomass waste residues in Tanzania," *Energies*, vol. 14, no. 6, p. 1793, 2021.

[16] A. Anderson and B. Rezaie, "Geothermal technology: trends and potential role in a sustainable future," *Applied Energy*, vol. 248, pp. 18–34, 2019.

[17] A. Carotenuto, G. Luca, S. Fabozzi et al., "Energy analysis of a small geothermal district heating system in Southern Italy," *International Journal of Heat and Technology*, vol. 34, no. Special Issue 2, pp. S519–S527, 2016.

[18] P. Yu, D. Dempsey, and R. Archer, "Techno-Economic feasibility of enhanced geothermal systems (EGS) with partially bridging Multi-Stage fractures for district heating applications," *Energy Conversion and Management*, vol. 257, p. 115405, 2022.

[19] M. Iorio, A. Carotenuto, A. Corniello et al., "Low enthalpy geothermal systems in structural controlled areas: a sustainability analysis of geothermal resource for heating plant (the Mondragone case in Southern Appennines, Italy)," *Energies*, vol. 13, no. 5, p. 1237, 2020.

[20] C. S. Brown, N. J. Cassidy, S. S. Egan, and D. Griffiths, "Thermal and economic analysis of heat exchangers as part of a geothermal district heating scheme in the cheshire basin, UK," *Energies*, vol. 15, no. 6, p. 1983, 2022.

[21] A. Ahmadi, M. El Haj Assad, D. H. Jamali et al., "Applications of geothermal organic Rankine Cycle for electricity production," *Journal of Cleaner Production*, vol. 274, p. 122950, 2020.

[22] J. M. Ngowi, L. Bångens, and E. O. Ahlgren, "Benefits and challenges to productive use of off-grid rural electrification: the case of mini-hydropower in Bulongwa-Tanzania," *Energy for Sustainable Development*, vol. 53, pp. 97–103, 2019.

[23] B. Kichonge, "The status and future prospects of hydropower for sustainable water and energy development in Tanzania," *Journal of Renewable Energy*, vol. 2018, pp. 1–12, 2018.

[24] A. Aly, S. S. Jensen, and A. B. Pedersen, "Solar power potential of Tanzania: identifying CSP and PV hot spots through a GIS multicriteria decision making analysis," *Renewable Energy*, vol. 113, pp. 159–175, 2017.

[25] A. A. Mas'ud and A. V. Wirba, "Muhammad-Sukki, F: a review on the recent progress made on solar photovoltaic in selected countries of sub-Saharan Africa," *Renewable and Sustainable Energy Reviews*, vol. 62, pp. 441–452, 2016.

[26] S. W. Chisale and Z. Sari, "Design of stand-alone solar-wind-hydro based hybrid power system: case of rural village in Malawi," *Journal of Energy Research and Reviews*, pp. 1–16, 2019.

[27] R. Poudel, "Performance and economic analysis of hybrid micro hydro systems," 2018.

[28] R. Syahputra and I. Soesanti, "Planning of hybrid micro-hydro and solar photovoltaic systems for rural areas of central Java, Indonesia," *Journal of Electrical and Computer Engineering*, vol. 2020, pp. 1–16, Article ID 5972342, 2020.

[29] M. D. A. Al-Falahi, S. D. G. Jayasinghe, and H. Enshaei, "A review on recent size optimization methodologies for standalone solar and wind hybrid renewable energy system," *Energy Conversion and Management*, vol. 143, pp. 252–274, 2017.

[30] F. A. Khan, N. Pal, and S. H. Saeed, "Review of solar photovoltaic and wind hybrid energy systems for sizing strategies optimization techniques and cost analysis methodologies," *Renewable and Sustainable Energy Reviews*, vol. 92, pp. 937–947, 2018.

[31] S. M. Zahraee, M. Khalaji Assadi, and R. Saidur, "Application of artificial intelligence methods for hybrid energy system optimization," *Renewable and Sustainable Energy Reviews*, vol. 66, pp. 617–630, 2016.

[32] U. K. Kalla, B. Singh, and S. S. Murthy, "Slide mode control of microgrid using small hydro driven single-phase SEIG integrated with solar PV array," *IET Renewable Power Generation*, vol. 11, no. 11, pp. 1464–1472, 2017.

[33] T. Li, W. Hu, X. Xu et al., "Optimized operation of hybrid system integrated with MHP, PV and PHS considering generation/load similarity," *IEEE Access*, vol. 7, pp. 107793–107804, 2019.

[34] M. J. Rajadurai, M. Shaanmaadhuran, J. V. Amaladoss, S. G. Sankar, and V. Kirubakaran, "Integration of solar PV—micro hydropower system for household application," in *Proceedings of the IEEE international conference on advances in electrical technology for green energy (ICAETGT)*, pp. 110–113, Coimbatore, India, September 2017.

[35] R. Syahputra and I. Soesanti, "Renewable energy systems based on micro-hydro and solar photovoltaic for rural areas: a case study in Yogyakarta, Indonesia," *Energy Reports*, vol. 7, pp. 472–490, 2021.

[36] R. C. Poudel, J. F. Manwell, and J. G. McGowan, "Performance analysis of hybrid microhydro power systems," *Energy Conversion and Management*, vol. 215, p. 112873, 2020.

[37] H. S. Das, A. H. M. Yatim, C. W. Tan, and K. Y. Lau, "Proposition of a PV/tidal powered micro-hydro and diesel hybrid system: a southern Bangladesh focus," *Renewable and Sustainable Energy Reviews*, vol. 53, pp. 1137–1148, 2016.

[38] W. Clements, K. Silwal, S. Pandit et al., "Unlocking electric cooking on Nepali micro-hydropower mini-grids," *Energy for Sustainable Development*, vol. 57, pp. 119–131, 2020.

[39] S. Hoseinzadeh, M. H. Ghasemi, and S. Heyns, "Application of hybrid systems in solution of low power generation at hot seasons for micro hydro systems," *Renewable Energy*, vol. 160, pp. 323–332, 2020.

[40] M. Das, M. A. K. Singh, and A. Biswas, "Techno-economic optimization of an off-grid hybrid renewable energy system using metaheuristic optimization approaches - case of a radio transmitter station in India," *Energy Conversion and Management*, vol. 185, pp. 339–352, 2019.

[41] H. Patel and S. Chowdhury, "Review of technical and economic challenges for implementing rural microgrids in South

Africa," in *Proceedings of the 2015 IEEE Eindhoven PowerTech*, pp. 1–6, Eindhoven, Netherlands, June 2015.

[42] D. A. Zema, A. Nicotra, V. Tamburino, and S. M. Zimbone, "A simple method to evaluate the technical and economic feasibility of micro hydro power plants in existing irrigation systems," *Renewable Energy*, vol. 85, pp. 498–506, 2016.

[43] L. M. Halabi, S. Mekhilef, L. Olatomiwa, and J. Hazelton, "Performance analysis of hybrid PV/diesel/battery system using HOMER: a case study Sabah, Malaysia," *Energy Conversion and Management*, vol. 144, pp. 322–339, 2017.

[44] M. Hossain, S. Mekhilef, and L. Olatomiwa, "Performance evaluation of a stand-alone PV-wind-diesel-battery hybrid system feasible for a large resort center in South China Sea, Malaysia," *Sustainable Cities and Society*, vol. 28, pp. 358–366, 2017.

[45] L. Olatomiwa, S. Mekhilef, and O. S. Ohunakin, "Hybrid renewable power supply for rural health clinics (RHC) in six geo-political zones of Nigeria," *Sustainable Energy Technologies and Assessments*, vol. 13, pp. 1–12, 2016.

[46] L. Olatomiwa, R. Blanchard, S. Mekhilef, and D. Akinyele, "Hybrid renewable energy supply for rural healthcare facilities: an approach to quality healthcare delivery," *Sustainable Energy Technologies and Assessments*, vol. 30, pp. 121–138, 2018.

[47] S. S. Yusuf and N. N. Mustapha, "Design and simulation of an optimal mini-grid solar-diesel hybrid power generation system in a remote Bangladesh," *International Journal of Smart Grids, ijSmartGrid*, vol. 2, no. 1, pp. 27–33, 2018.

[48] N. Yimen, T. Tchotang, A. Kanmogne et al., "Optimal sizing and techno-economic analysis of hybrid renewable energy systems-A case study of a photovoltaic/wind/battery/diesel system in fanisau, northern Nigeria," *Processes*, vol. 8, no. 11, p. 1381, 2020.

[49] M. J. Hadidian-Moghaddam, S. Arabi-Nowdeh, and M. Bigdeli, "Optimal sizing of a stand-alone hybrid photovoltaic/wind system using new grey wolf optimizer considering reliability," *Journal of Renewable and Sustainable Energy*, vol. 8, no. 3, p. 035903, 2016.

[50] A. Aly, M. Moner-Girona, S. Szabó, and A. B. S. S. Pedersen, "Barriers to large-scale solar power in Tanzania," *Energy for Sustainable Development*, vol. 48, pp. 43–58, 2019.

[51] O. J. Mdee, C. Z. Kimambo, T. K. Nielsen, and J. Kihedu, "Measurement methods for hydropower resources: a review," *Water Utility Journal*, vol. 18, pp. 21–38, 2018.

[52] D. K. Geleta, M. S. Manshahia, P. Vasant, and A. Banik, "Grey wolf optimizer for optimal sizing of hybrid wind and solar renewable energy system," *Computational Intelligence*, 2020.

[53] A. Aly, A. Bernardos, C. M. Fernandez-Peruchena, and S. S. A. B. Jensen, "Is concentrated solar power (CSP) a feasible option for sub-saharan africa?: i," *Renewable Energy*, vol. 135, pp. 1224–1240, 2019.

[54] M. Combe, A. Mahmoudi, M. H. Haque, and R. Khezri, "Cost-effective sizing of an AC mini-grid hybrid power system for a remote area in South Australia," *IET Generation, Transmission & Distribution*, vol. 13, no. 2, pp. 277–287, 2019.

[55] NASA, "NASA surface meteorology and solar energy retrieved 29, April," 2019, http://www.nasa.gov.

[56] A. Kumar, M. Rizwan, and U. Nangia, "A new approach to design and optimize sizing of hybrid micro grid in deregulated electricity environment," *CSEE Journal of Power and Energy Systems*, vol. 8, 2020.

[57] M. A. M. Shaheen, H. M. Hasanien, and A. Alkuhayli, "A novel hybrid GWO-PSO optimization technique for optimal reactive power dispatch problem solution," *Ain Shams Engineering Journal*, vol. 12, no. 1, pp. 621–630, 2021.

[58] N. Chopra, G. Kumar, and S. Mehta, "Hybrid GWO-PSO algorithm for solving convex economic load dispatch problem," *Int J Res Adv Technol*, vol. 4, no. 6, pp. 37–41, 2016.

[59] K. Sakulphan and E. Bohez, "A new optimal selection method with seasonal flow and irrigation variability for hydro turbine type and size," *Energies*, vol. 11, no. 11, p. 3212, 2018.

[60] P. Anand, M. Rizwan, and S. K. Bath, "Sizing of renewable energy based hybrid system for rural electrification using grey wolf optimisation approach," *IET Energy Systems Integration*, vol. 1, no. 3, pp. 158–172, 2019.

[61] K. Sharma, S. Kothari, N. L. Panwar, and N. Rathore, "Design and development of solar energy powered maize milling machine International," *Journal of Ambient Energy*, pp. 1–6, 2020.

[62] I. Sansa, R. Villafafila, and N. M. Bellaaj, "Optimal sizing design of an isolated microgrid using loss of power supply probability," in *Proceedings of the IREC2015 The Sixth International Renewable Energy Congress*, pp. 1–7, Sousse, Tunisia, March 2015.

[63] N. I. Abdul Aziz, S. I. Sulaiman, S. Shaari, I. Musirin, and K. Sopian, "Optimal sizing of stand-alone photovoltaic system by minimizing the loss of power supply probability," *Solar Energy*, vol. 150, pp. 220–228, 2017.

[64] D. Borkowski, D. Cholewa, and A. Korzeń, "Run-of-the-River hydro-PV battery hybrid system as an energy supplier for local loads," *Energies*, vol. 14, no. 16, p. 5160, 2021.

[65] A. Mardani, A. Jusoh, E. Zavadskas, F. Cavallaro, and Z. Khalifah, "Sustainable and renewable energy: an overview of the application of multiple criteria decision making techniques and approaches," *Sustainability*, vol. 7, no. 10, pp. 13947–13984, 2015.

[66] I. Siksnelyte-Butkiene, E. K. Zavadskas, and D. Streimikiene, "Multi-criteria decision-making (MCDM) for the assessment of renewable energy technologies in a household: a review," *Energies*, vol. 13, no. 5, p. 1164, 2020.

[67] O. Drakenberg, G. Ek, and K. W. Fernqvist, "Environmental and Climate. Change Policy brief—Tanzania," *Sidas technical report for Sida helpdesk for environment and climate change*, 2016.

[68] A. K. Othman, O. Wan, and G. Jamel, "Water flow measuring methods in small hydropower for streams and rivers-A study," *International Journal of Applied Engineering Research*, vol. 12, pp. 14484–14489, 2017.

[69] H. Bakhtiari and R. A. Naghizadeh, "Multi-criteria optimal sizing of hybrid renewable energy systems including wind, photovoltaic, battery, and hydrogen storage with ε-constraint method," *IET Renewable Power Generation*, vol. 12, no. 8, pp. 883–892, 2018.

[70] Tanzania Meteorological Agency (TMA) Data Section, "Mbeya meteorological station (MMS)," 2019.

PeerJ Computer Science

Predicting defects in imbalanced data using resampling methods: an empirical investigation

Ruchika Malhotra[1] and Juhi Jain[2]

[1] Department of Software Engineering, Delhi Technological University (former Delhi College of Engineering), Shahbad Daulatpur, Delhi, India
[2] Department of Computer Science and Engineering, Delhi Technological University (former Delhi College of Engineering), Shahbad Daulatpur, Delhi, India

ABSTRACT

The development of correct and effective software defect prediction (SDP) models is one of the utmost needs of the software industry. Statistics of many defect-related open-source data sets depict the class imbalance problem in object-oriented projects. Models trained on imbalanced data leads to inaccurate future predictions owing to biased learning and ineffective defect prediction. In addition to this large number of software metrics degrades the model performance. This study aims at (1) identification of useful metrics in the software using correlation feature selection, (2) extensive comparative analysis of 10 resampling methods to generate effective machine learning models for imbalanced data, (3) inclusion of stable performance evaluators—AUC, GMean, and Balance and (4) integration of statistical validation of results. The impact of 10 resampling methods is analyzed on selected features of 12 object-oriented Apache datasets using 15 machine learning techniques.
The performances of developed models are analyzed using AUC, GMean, Balance, and sensitivity. Statistical results advocate the use of resampling methods to improve SDP. Random oversampling portrays the best predictive capability of developed defect prediction models. The study provides a guideline for identifying metrics that are influential for SDP. The performances of oversampling methods are superior to undersampling methods.

Submitted 30 October 2020
Accepted 11 May 2021
Published 29 April 2022

Corresponding author
Juhi Jain,
juhijain_phdco2k16@dtu.ac.in

Academic editor
Stefan Wagner

Additional Information and Declarations can be found on page 29

DOI **10.7717/peerj-cs.573**

Copyright
2022 Malhotra and Jain

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

Subjects Data Mining and Machine Learning, Software Engineering
Keywords Software defect prediction, Machine learning, Class imbalance problem, Resampling methods, Statistical validation

INTRODUCTION

Software Defect Prediction (SDP) deals with uncovering the probable future defects. Efficient defect prediction helps in the timely identification of areas in software that can lead to defects in software owing to better resource utilization (*Malhotra, 2016*). Source code metrics give useful insights to software quality attributes like cohesion, coupling, size, inheritance, etc, and are extensively used in developing software defect models (*Basili, Briand & Melo, 1996*; *Singh, Kaur & Malhotra, 2010*; *Radjenović et al., 2013*). Effective models can be generated using the object-oriented (OO) metrics. With the increasing

complexity of software, early prediction of defects, and assurance of good software quality of projects become difficult tasks. To achieve this task of SDP, machine learning (ML) techniques have been used by several researchers from the past two decades. But software defect data is mostly imbalanced (*Khoshgoftaar, Gao & Seliya, 2010*). This issue of SDP has recently gained a high interest in researchers in the software engineering community. Data is imbalanced if the number of minority classes, in our case, defective classes, is much lower than the majority class, i.e., non-defective classes. This imbalanced distribution of data misguides classifiers while learning the defective class correctly and hence results in biased and inaccurate results. A good defect prediction model will be the one that is trained on similar distribution of instances of defective and non-defective classes. From this point onwards, we will refer to the imbalanced data problem as the class imbalance problem (CIP) because we are dealing with an imbalanced ratio of defective and non-defective classes of software.

CIP is very prominent in many real-life problems like medical diagnosis (*Mazurowski et al., 2008*; *Khalilia, Chakraborty & Popescu, 2011*), fraud detection (*Phua, Alahakoon & Lee, 2004*; *Vasu & Ravi, 2011*), text categorization (*Zheng, Wu & Srihari, 2004*; *Moreo, Esuli & Sebastiani, 2016*), sentiment analysis (*Lane, Clarke & Hender, 2012*), churn prediction (*Burez & Van den Poel, 2009*), *etc*. Researchers are actively participating in finding solutions for this problem in the software engineering domain for the last two decades. One of the prominent solutions is using resampling methods comprising of undersampling and oversampling methods (*Kotsiantis, Kanellopoulos & Pintelas, 2006*). A large number of software metrics adds to the problem of constructing good SDP models. This curse of dimensionality is handled by selecting important and distinct features using correlation feature selection (CFS). CFS is selected as it is a widely accepted feature selection (FS) technique and it has emerged as an effective FS technique in a benchmark study (*Ghotra, McIntosh & Hassan, 2017*). This study deals with the application of resampling methods to handle CIP for 12 Apache datasets and performing a comparative analysis of various SDP models developed using ML techniques. In this study, six oversampling techniques namely Synthetic Minority Over-sampling Technique (SMT), Safe Level Synthetic Minority Over-sampling Technique (SLSMT), Selective Preprocessing of Imbalanced Data (SPD), ADAptive SYNthetic Sampling (ADSYN), Random OverSampling (ROS), Aglomerative Hierarchical Clustering (AHC) and four undersampling techniques namely Condensed Nearest Neighbor + Tomek's modification of Condensed Nearest Neighbor (CNNTL), Random Under-Sampling (RUS), Neighborhood Cleaning Rule (NCL), One Sided Selection (OSS) are explored on different datasets to deal with their imbalanced nature.

The main objective of this paper is to ascertain the importance of dealing with CIP using (1) resampling methods and (2) stable performance evaluators to build correct and effective SDP models. The study also identifies the important internal quality attributes (IQAs) of software on which developers need to focus.

Research Questions (RQs) to achieve the aforementioned objectives are designed as follows.

RQ1: Which features are repeatedly selected by CFS in software engineering datasets?

RQ1 finds the most important internal metrics that impact the possibility of defect(s) in the software. Software metrics are recognized that software designers and developers should focus on while building the software. We found that defects in software engineering datasets are greatly impacted by LCOM, Ca, Ce, and RFC. LCOM is a cohesion metric whereas Ca, Ce, and RFC are coupling metrics. In any software, high cohesion and low coupling are desired. Therefore, as LCOM, Ca, Ce and RFC play a crucial role in SDP, their values should be monitored while designing the software. This would result in software with fewer defects.

RQ2: What is the performance of ML techniques on imbalanced data while building SDP models?

RQ2 summarizes the performance of 12 imbalanced datasets with 15 ML techniques in terms of Sensitivity, GMean, Balance, and AUC.

RQ3a: What is the comparative performance of various SDP models developed using resampling methods?

RQ3b: Is there any improvement in the performance of SDP models developed using resampling methods?

RQ3 presents the performance of SDP models that are built on balanced datasets. The ratio of defective and non-defective classes is equalized with help of different resampling methods and their performance is compared with their original versions, i.e. when no resampling of classes is done. We empirically proved that all resampling methods improve the performance of SDP models as compared to the models that are trained on original data. The values of AUC, Balance, GMean, and Sensitivity increased with models trained on resampled data, and, hence, these models depict better defect prediction capabilities.

RQ4: Which resampling method outperforms the addressed undersampling and oversampling methods for building an efficient SDP model?

In RQ3 it is proved that resampling methods produce better SDP models but out of the wide array of resampling methods which resampling method should the developer or researcher choose? RQ4 is generated to find out the resampling method(s) that outperforms the other resampling methods. The results of the Friedman Test followed by the post-hoc Nemenyi Test obtained on AUC, Balance, GMean, and Sensitivity values show that ML models based on ROS and AHC statistically outperforms other models.

RQ5: Which ML technique performs the best for SDP in imbalanced data?

This study uses 15 ML techniques from different categories. RQ5 tries to explore if there is a statistical difference in the performances of different ML techniques. We constricted the comparison to the models developed using ROS as it was statistically better than other resampling methods except for AHC. Performance values of ROS-based models were comparable but greater than the performance values of AHC-based models. We concluded that statistically ensemble methods and nearest neighbor methods performed better than neural networks and statistical techniques.

The answers to these questions are explored by building ML models on CFS selected features using ten-fold cross-validation. Predictive performances of developed models are evaluated using stable performance evaluators like Sensitivity, GMean, Balance, and AUC. *Kitchenham et al. (2017)* reported that in many studies results are biased because they lack statistical validation. They recommend using a robust statistical test to examine if performance differences are significant or not. Statistical validation is carried out using the Friedman test followed by post hoc analysis that is performed using the Nemenyi test. The conducted study will acquaint developers with useful resampling methods and performance evaluators that will assist them to solve CIP. This study also guides developers and software practitioners about the important metrics that affect the SDP potential of ML models. The result examination ascertained ROS-based and AHC-based ML models as the best defect predictors for datasets related to the software engineering domain. With ROS as a resampling method, nearest neighbors and ensembles exhibited comparable performance in SDP. These models were statistically better than other ML models.

The rest of the paper is organized as follows. "Related work" deals with research work done in SDP for imbalanced data. Empirical Study Design is explained in "Materials and Methods". "Results" expounds on the empirical findings and provides answers to set RQs. Next, "Discussions" summarizes the results and provides the comparison of this study with related studies. "Validity Threats" uncovers the validity threats of this study. Finally, "Conclusions" presents the concluding remarks with potential future directions.

RELATED WORK

This section presents the related work done in the field of feature selection and resampling solutions proposed in the SDP field.

Feature selection in SDP

Apache datasets have twenty OO metrics and models developed using all these metrics can hamper the defect prediction capabilities of ML models. The reason is the presence of redundant or irrelevant metrics. This curse of dimensionality can be reduced by using feature reduction strategies. This involves either feature selection—reducing the number of features, or feature extraction—extracting new features from existing ones. This study focuses on feature selection by using a widely acceptable CFS technique. *Ghotra, McIntosh & Hassan (2017)* explored 30 feature selection techniques and concluded CFS as the best feature predictor. They used NASA datasets and PROMISE datasets with 21 ML

techniques. *Balogun et al. (2019)* explored feature selection and feature reduction methods for five NASA datasets over four ML techniques and experimentally concluded that FS techniques did not show consistent behavior for the datasets or ML techniques. Recent studies (*Arar & Ayan, 2017*; *Lingden et al., 2019*) have emphasized the importance of feature selection and the impact of CFS in building efficient models with reduced complexity and computation time. *Balogun et al. (2020)* empirically investigated the effect of 46 FS methods over 25 datasets from different sources using Naïve Bayes and decision trees. Based on accuracy and AUC performance, they concluded CFS was the best performer in the FSS category. Therefore, in this study, CFS is used to reveal the most relevant features.

Solutions proposed for CIP in SDP

Only 20% of software classes are accountable for the defects in software (*Koru & Tian, 2005*). This principle is enough to explain the reason for the uneven distribution of minority (defective) classes and majority (non-defective) classes. Areas of SDP and software change prediction (*Malhotra & Khanna, 2017*; *Tan et al., 2015*) are explored to handle CIP resulting in promising outcomes. Now, to solve this class imbalance issue, a variety of resampling methods have been proposed in the literature, among which oversampling and undersampling techniques are most widely used. *Liu, An & Huang (2006)* used a combination of oversampling and undersampling techniques for predicting software defects using a support vector machine. The performance of developed models was evaluated using F-Measure, GMean, and ROC curve. Experimentation by *Pelayo & Dick (2007)* showed improvement in GMean values for SDP when the oversampling technique SMT is used with a C4.5 decision tree classifier. *Kamei et al. (2007)* evaluated the effect of ROS, RUS, SMT, and OSS on industrial software. The experimental analysis proved that resampling methods improved the performance of LDA and LR models in terms of F1-measure. *Khoshgoftaar & Gao (2009)* used RUS to handle CIP and also used a wrapper-based feature selection technique for attribute selection. They investigated four different scenarios of sampling techniques and feature selection combinations to evaluate which model has better predictive capability in terms of accuracy and AUC.

Galar et al. (2011) performed SDP for imbalanced data using bagging—and boosting—based ensemble techniques with C4.5 as the base classifier. The performance was evaluated using the AUC measure. Some other studies support the application of resampling methods for handling CIP (*Riquelme et al., 2008*; *Pelayo & Dick, 2012*; *Seiffert et al., 2014*; *Rodriguez et al., 2014*). *Shatnawi (2012)* also performed an empirical comparison of defect prediction models built using oversampling techniques with three different classifiers on the eclipse dataset. *Wang & Yao (2013)* investigated ML models built using five different resampling methods on 10 PROMISE datasets and their findings confirmed that models based on resampled data result in better SDP. Experimentations concluded the effective model development with AdaboostNC ensemble. *Wang & Yao (2013)* were not able to conclude on which resampling method should be selected by the software practitioners. We solved this issue by providing detailed statistical analysis

and experimentally proved ROS and AHC to be the better options for researchers and other practitioners.

Jindaluang, Chouvatut & Kantabutra (2014) proposed the undersampling technique with the k-centers clustering algorithm which proves to be effective in terms of Sensitivity and F-measure, but they didn't use any stable metric for imbalanced data like GMean or AUC.

Bennin et al. (2016) concluded that ROS outperformed the SMT approach when defective classes are less than 20% in the software. This study also empirically proved that ROS is statistically better than SMT, hence supports the conclusions of *Bennin et al. (2016)*.

Tantithamthavorn, Hassan & Matsumoto (2018) included a large number of datasets but provided a comparative study with only four resampling methods. They also optimized SMT and found its performance comparable with RUS. We created 1980 models with 12 datasets as compared to their 4242 models with 101 datasets. In this study, we covered 10 resampling methods and provided the breadth of resampling methods. We compared models with 15 ML techniques whereas *Tantithamthavorn, Hassan & Matsumoto (2018)* used only 7 ML techniques. The results of our study are different from those of *Tantithamthavorn, Hassan & Matsumoto (2018)*. The main reasons are that the model evaluation is highly dependent on the nature of the data and the behavior of ML techniques. This problem amplifies by adding the third dimension of resampling methods. Furthermore, *Bennin et al. (2018)* empirically concluded that AUC performance is not improved by the SMT and this contradicts with the results of *Tantithamthavorn, Hassan & Matsumoto (2018)*.

Agrawal & Menzies (2018) also proposed the modified SMT by tuning the SMT parameters. They emphasized on the fact that preprocessing, i.e., resampling technique is more important factor than the ML technique used for the construction of prediction model. If data could be better (less skewed), then results would be more reliable. *Malhotra & Kamal (2019)* inspected the impact of oversampling techniques on ML models built with 12 NASA datasets. They demonstrated the improvement in ML models with oversampling and proposed a new resampling method—SPIDER3.

Though many studies have been conducted, still there is no particular set of resampling methods that can be considered the winner of all. These techniques certainly need more replicated studies with different classifiers and different datasets. More often NASA datasets are exploited by researchers for investigating CIP. We have comparatively used Apache datasets to visualize the effect of sampling techniques. Repetitive studies are required to be performed in the future for a fair comparison. Apart from decision tree-based and ensemble-based classifiers, this study used rule-based, neural network-based, and statistical-based machine learners for assessing the predictive capability of models.

MATERIALS & METHODS

This section describes the components involved in this empirical study. This section describes the framework established to build a classification model for defect prediction

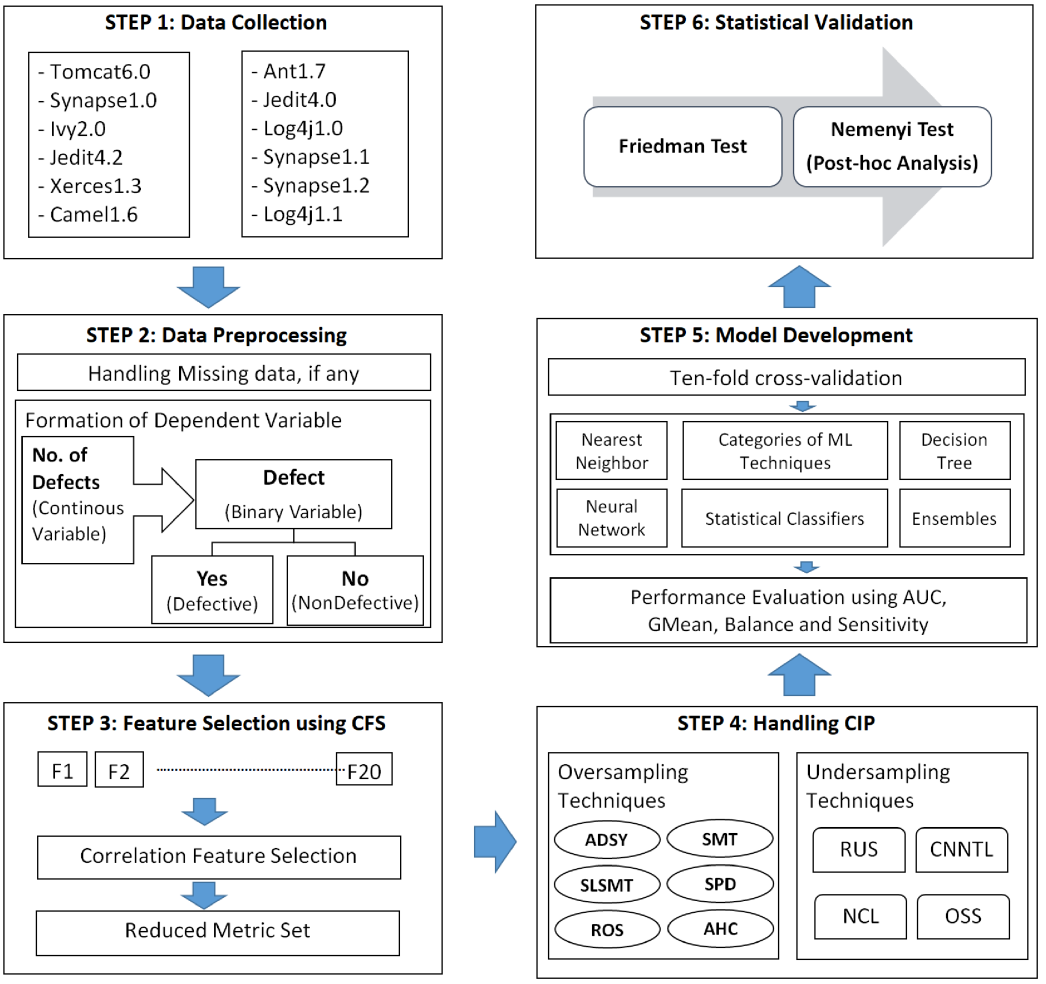


Figure 1 **Experimental framework for SDP in imbalanced data.**
Full-size DOI: 10.7717/peerj-cs.573/fig-1

from dataset collection to model validation. Figure 1 explains the experimental setup for the study.

Step1: dataset collection

Datasets mined from the PROMISE repository were used for empirical predictive modeling and validation. These datasets were collected by *Jureczko & Spinellis (2010)*. *Jureczko & Madeyski (2010)* used these datasets to find clusters of similar software projects. The datasets consisted of 20 software metrics and a dependent variable indicating the number of defects in a particular class. The selection of projects was based on the percentage of data imbalance in them. Details of the datasets are provided in Table 1. #Classes denotes the total number of classes in the project, #DClasses denotes the number of defective classes, and #%ageDefects represents the percentage of defective classes in a project. The percentage of defective classes in addressed projects varies from 9.85% to 33.9%. This low percentage represents the imbalanced ratio of defective and non-defective classes in the datasets.

Table 1 Dataset details. #Classes denotes the total number of classes in the project, #DClasses denotes the number of defective classes, and #%ageDefects represents the percentage of defective classes in a project.

Dataset	#Classes	#DClasses	#%ageDefects
Tomcat6.0	858	77	9.85
Synapse1.0	157	16	10.19
Ivy2.0	352	40	11.4
Jedit4.2	367	48	13.1
Xerces1.3	453	69	15.2
Camel1.6	965	188	19.48
Ant1.7	745	166	22.3
Jedit4.0	306	75	24.5
Log4j1.0	135	34	25.2
Synapse1.1	222	60	27
Synapse1.2	256	86	33.6
Log4j1.1	109	37	33.9

Step 2: data preprocessing and identification of variables

The independent variables used in this study are different OO metrics characterizing a software system from various aspects. The OO metrics used in the study include the following metrics:

***Chidamber and Kemerer (CK) Metric suite* (*Chidamber & Kemerer, 1994*):** Six popularly used metrics, namely Weighted Methods of a Class (WMC), Depth of Inheritance Tree (DIT), Number of Children (NOC), Lack of Cohesion in Methods (LCOM), Response For a Class (RFC) and Coupling Between Objects (CBO), are incorporated in this metric suite. This metric suite has been validated in many empirical studies for developing SDP models.

***Quality Model for Object-Oriented Design metric suite (QMOOD)* (*Bansiya & Davis, 2002*):** The study uses few metrics from this metric suite namely Number of Public Methods (NPM), Data Access Metric (DAM), Method of Functional Abstraction (MFA), Measure of Aggression (MOA) and Cohesion among Methods of a class (CAM) along with CK metrics for developing defect prediction models. These metrics are also well exploited in related studies to develop effective software quality prediction models.

***Other metrics*:** Few other metrics have also been used in this study as independent variables in addition to the above metrics that are widely used by researchers. These metrics are—Efferent Coupling (Ce), Afferent Coupling (Ca), Lines of Code (LOC), Coupling Between Methods of a Class (CBM), Average Method Complexity (AMC), and the variant of LCOM (LCOM3). Ce and Ca are proposed by *Martin (1994)* and LOC, CBM, AMC, and LCOM3 are proposed by *Henderson-Sellers (1995)*. Inheritance Coupling (IC), maximum Cyclomatic complexity (Max_cc), and average cyclomatic complexity (Avg_cc) are also used in addition to the above metrics.

Details of metrics used can be referred to from http://gromit.iiar.pwr.wroc.pl/p_inf/ckjm/metric.html.

Table 2 OO features selected by CFS.

Dataset	Features selected by CFS
Tomcat6.0	CBO, RFC, LCOM, LOC, MOA, AMC, Max_CC, Avg_CC
Synapse1.0	RFC, LCOM, Ce, LOC, DAM, MFA, CAM, IC, AMC, Max_CC, Avg_CC
Ivy2.0	RFC, Ce, LOC, MOA, AMC, NPM, CBO, WMC, LCOM, LCOM3, CAM
Jedit4.2	CBO, RFC, Ca, Ce, NPM, LCOM3, MOA, LOC, CAM, CBM, LCOM, AMC, Max_CC
Xerces1.3	WMC, Ce, LCOM, MOA, DAM, IC, CBM, AMC
Camel1.6	DIT, NOC, CBO, LCOM, Ca, NPM, LCOM3, CAM, IC, CBM, AMC, Max_CC, Avg_CC
Ant1.7	CBO, RFC, LCOM, Ce, LOC, MOA, CAM, AMC, Max_CC
Jedit4.0	WMC, DIT, CBO, RFC, LCOM, Ca, Ce, NPM, LCOM3, LOC, DAM, MOA, CBM, Max_CC, Avg_CC
Log4j1.0	WMC, CBO, RFC, LCOM, Ce, Ca, NPM, LOC, DAM, CAM, Avg_CC
Synapse1.1	DIT, CBO, RFC, LCOM, Ce, DAM, MFA, CAM, Max_CC
Synapse1.2	WMC, CBO, RFC, Ca, Ce, LOC, MOA, CBM, CAM, AMC, Max_CC
Log4j1.1	WMC, RFC, LCOM, Ce, NPM, LCOM3, MOA, MFA

Datasets are checked for any inconsistencies like missing data or redundant data. If there are data inconsistencies, the model prediction would be biased. Therefore, it was important to clean the data before using it for model development. Datasets collected contain a continuous variable representing the number of defects. It was converted into a binary variable by replacing '0' with 'No' and natural numbers with 'yes'. The binary dependent variable, 'defect' with two possible values are 'yes' and 'no' reflects whether the software class is defective or not.

Step 3: feature selection

All these metrics may not be important in the concern of predicting defects in the early stages of project development. This study employs CFS for the identification of significant metrics. A review by *Malhotra (2015)* has revealed that CFS is the most commonly used feature selection technique. CFS is used in this study for selecting features because it is the most preferred FS technique in SDP literature (*Ghotra, McIntosh & Hassan, 2017*; *Arar & Ayan, 2017*; *Lingden et al., 2019*; *Balogun et al., 2020*). CFS performs the ranking of features based on information gain and identifies the optimal subset of features. The features in the subset are highly correlated to the class label 'defect' and are uncorrelated or less correlated with each other. CFS is incorporated to minimize the multicollinearity effect.

A list of metrics selected by CFS for each dataset is presented in Table 2.

Step 4: handling CIP

This step involves applying resampling methods for treating CIP. These methods are implemented using a knowledge extraction tool based on evolutionary learning (KEEL) (*Alcalá-Fdez et al., 2011*). Six oversampling methods and four undersampling methods were applied to create a balance between majority and minority classes. Details of these resampling methods are given in Table 3.

Table 3 Resampling methods used in the study.

Resampling Methods		Description
Oversampling Methods	ADAptive SYNthetic Sampling (ADSYN) (*He et al., 2008*)	In ADASYN, synthetic samples are generated by finding the density distribution of minority classes. Density distribution is computed using a k-nearest neighbor with Euclidian distance. It is an extension of the Synthetic Minority Oversampling Technique. It focuses on the samples that are hard to classify.
	Synthetic Minority Over-sampling Technique (SMT) (*Chawla et al., 2002*)	The number of minority class samples is increased by generating artificial samples in direction of k nearest neighbors of minority class samples. If one neighbor is selected, then one synthetic sample is generated corresponding to that original minority sample resulting in 100% oversampling of minority classes. In this study, k = 5 results in 500% oversampling of minority classes.
	Safe Level Synthetic Minority Oversampling Technique (SLSMT) (*Bunkhumpornpat, Sinapiromsaran & Lursinsap, 2009*)	Unlike the SMT version, where synthetic samples are generated randomly, in SLSMT first safe levels are calculated that helps in determining safe positions for generating the synthetic samples. If the safe level value is close to 0, it is considered noise and if the safe level value is close to k for k nearest neighbor implementation, then it is considered safe resulting in producing synthetic minority samples there.
	Selective Preprocessing of Imbalanced Data (SPD) (*Stefanowski & Wilk, 2008*)	This technique categorizes the sample as either safe or noise based on the nearest neighbor rule where distance measurement is done using a heterogeneous value distance metric. If an instance is accurately classified by its k nearest neighbors, it is considered safe otherwise it is considered noise and then discarded.
	Random OverSampling (ROS) (*Batista, Prati & Monard, 2004*)	ROS is a very simple oversampling technique in which minority class instances are replicated at random with the sole aim of creating a balance between majority and minority class instances.
	Agglomerative Hierarchical Clustering (AHC) (*Cohen et al., 2006*)	In AHC, each class is decomposed into sub-clusters and synthetic samples are generated corresponding to cluster prototypes. Since artificial samples are created as centroids of sub-clusters of classes, they, therefore extract the characteristics of that class and represent better samples than randomly generated samples.
Undersampling Methods	Random UnderSampling (RUS) (*Batista, Prati & Monard, 2004*)	RUS, like ROS, is a non-heuristic technique. But in this instead of replicating minority class instances, majority class instances are removed with aim of creating a balance between majority and minority class instances. The problem with this technique is that some important or useful data may be rejected as it is based on random selection.
	Condensed Nearest Neighbor(CNN) + Tomek's modification of Condensed Nearest Neighbor (CNNTL) (*Cohen et al., 2006*)	First CNN is applied to find a consistent subset of samples that helps to eliminate majority class samples that are far from the decision border. Then Tomek links (*Tomek, 1976*) are made between samples. If it exists between any two samples, then either both are borderline samples or one of them is noise. Samples with Tomek links that fit in majority classes are removed.
	Neighborhood Cleaning Rule (NCL) (*Laurikkala, 2001*)	NCL uses the edited nearest-neighbor (ENN) rule to remove majority class samples. For each training sample Si, first it finds three nearest neighbors. If Si= majority class sample, then discard it if more than two nearest neighbors incorrectly classifies it. If Si = minority class sample, then discard the nearest neighbors if they incorrectly classify it.
	One Sided Selection (OSS) (*Kubat & Matwin, 1997*)	OSS and CNNTL have similar working. The difference lies in the order of the application of CNN and the determination of Tomek links. OSS identifies unsafe samples using Tomek links and then applying condensed nearest neighbor (CNN). Noisy and borderline samples are considered unsafe. Small noise may result in the flipping of the decision border of the borderline samples; therefore, they are also considered unsafe. CNN eliminates the majority of samples that are far away from decision boundaries.

Table 4 Confusion matrix.

	Predicted Positive	Predicted Negative
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

Step 5: performance evaluation and model development

Performance evaluators

When it comes to imbalanced data, the selection of appropriate performance evaluators plays a critical role. These measures can be calculated by using the confusion matrix shown in Table 4. TP represents the number of defective classes predicted correctly. TN represents the number of non-defective classes predicted correctly. FP represents the number of non-defective classes that are wrongly predicted as defective classes. FN represents the number of defective classes that are wrongly predicted as non-defective classes.

The use of accuracy to evaluate performance is specious when data is imbalanced. Instead, robust performance evaluators like AUC, GMean, and Balance should be used in the class imbalance framework. The Sensitivity indicates the probability of correctly predicted defective classes out of total defective classes. whereas specificity refers to the probability of identifying non-defective classes correctly. Sensitivity or True Positive Rate (TPR) is defined as

$$Sensitivity = \frac{TP}{TP + FN} \tag{1}$$

GMean maintains a balance between both these accuracies (*Li et al., 2012*). Therefore, it is wise to use GMean as an effective measure to assess imbalanced data. GMean is defined as the geometric mean of Sensitivity and specificity for any classifier.

$$GMean = \sqrt{sensitivity * specificity} \tag{2}$$

where

$$Specificity = \frac{TN}{TN + FP} \tag{3}$$

Balance corresponds to the Euclidean distance between a pair of Sensitivity and False Positive Rate (FPR) (*Li et al., 2012*). FPR is the probability of false alarm. It exemplifies the proportion of non-defective classes that are misclassified as defective classes amongst actual non-defective classes. Balance can be computed as—

$$Balance = 1 - \sqrt{\frac{\left(0 \quad \frac{FPR}{100}\right)^2 + \left(1 \quad \frac{Sensitivity}{100}\right)^2}{2}} \tag{4}$$

where

$$\text{FPR} = \frac{FP}{TN + FP} \tag{5}$$

The area under the curve (AUC) is widely accepted as a consistent and robust performance evaluator for predictions in imbalanced data (*Fawcett, 2006*; *Lessmann et al., 2008*; *Malhotra & Khanna, 2017*). It is threshold independent and can handle skewed data. It is a measure to distinguish between the two classes. The range of AUC is (0, 1). Higher the AUC value the better the prediction model. AUC value of 0.5 signifies that the model cannot differentiate between the two classes. AUC values from 0.7 to 0.8 are considered acceptable. AUC values greater than 0.8 are considered excellent.

Model development using ML techniques

This study developed models based on 15 ML techniques. Ten-fold within project cross-validation is carried out to reduce the partitioning bias. Data was divided into ten partitions. Nine partitions were used for the training part and the remaining one partition was used for the testing part. Then performance evaluators are averaged across ten folds. The ML parameters that were used in experiments created in this study are noted in Table 5. ML techniques used in this study can be divided into five major categories as described below.

Statistical techniques

- *Naive Bayes (NB)* (*John & Langley, 1995*): Naïve Bayes is a probability-based classifier that works on the Bayes theorem. It is an instance-based learner that computes class wise conditional probabilities. Features need to be conditionally independent with each other. It provides fair results even in violation of this assumption. This ML technique works well for both categorical and numerical variables.
- *Logistic Regression (LR)* (*Le Cessie & Van Houwelingen, 1992*): Logistic Regression is also a probabilistic classifier used for dichotomous variables and assumes that the data follows Gaussian distribution. It works well in case of assumption desecration. During training coefficient values are minimized by ridge estimator to solve multicollinearity and this makes the model simpler. The algorithm runs until it converged.
- *Simple Logistic (SL)* (*Sumner, Frank & Hall, 2005*): SimpleLogistic uses LogitBoost to construct logistic regression models. LogitBoost uses the logit transform to predict the probabilities. With each repetition, one simple regression model is added for each class. The process terminates when there is no more reduction in classification error.
- *LogitBoost (LB)* (*Friedman, Hastie & Tibshirani, 2000*): LogitBoost is an additive logistic regression with a decision stump as the base classifier. It maximizes the likelihood and, therefore, generalizes the linear logistic model. The base classifier taken is the decision stump which considers entropy for classification.

Neural networks

- *MultiLayerPerceptron (MLP)* (Rojas & Feldman, 2013): It is a backpropagation neural network that uses sigmoid function as the activation function. The number of hidden layers in the network is determined by the average of the number of attributes and total classes for a particular dataset. The error is backpropagated in every epoch and reduced via gradient descent. The network is then learned based on revised weights.

Nearest neighbors

- *IBk* (*Aha, Kibler & Albert, 1991*): IBk is an instance-based K-nearest neighbor learner. It calculates the Euclidian distance measure of the test sample with all the training samples to find its 'k' nearest neighbors. It then assigns the class label to the testing instance based on the majority classification of nearest neighbors. Only one nearest neighbor is determined with k = 1 and the class label of that nearest neighbor is assigned to the testing instance.
- *Kstar* (*Cleary & Trigg, 1995*): Like Ibk, Kstar is also an instance-based learning algorithm. The difference between the two techniques is about the similarity measures they use. IBk exploits Euclidian distance and Kstar uses the similarity measure based on entropy. Kstar exhibits good classification competence for noisy and imbalanced data.

Ensembles

- *AdaboostM1 (ABM1)* (*Quinlan, 1993*; *Freund & Schapire, 1996*): AaboostM1 is an ensemble technique where numbers of weak classifiers are used iteratively to improve the overall performance. It augments the performance of weak learners by adjusting the weak hypothesis returned by the weak learner. The base decision tree used in ABM1 is J48. J48 learns from the previous trees about misclassified instances and calculates the weighted average. NumIterations in parameters represent the number of classifiers involved in this ensemble. This technique helps in reducing the bias in the model.
- *Bagging (Bag)* (*Quinlan, 1993*; *Breiman, 1996*): Bagging or bootstrap aggregation is also one of the ensemble techniques that improve the predictive capability of base classifiers by making bags of training data. Models work in parallel and their results are averaged. Bagging reduces the variance. The number of bags used for experimentation is 10 and the base classifier used is J48.
- *Iterative Classifier Optimizer* (*ICO*): LogitBoost is used as the iterative classifier in this technique. Cross-validation is utilized for its optimization. In the experiments conducted, it goes through 50 iterations to decide for the best cross-validation.
- *Logistic Model Tree (LMT)* (*Landwehr, Hall & Frank, 2005*): Logistic Model Tree is a meta-learning algorithm that uses logistic regression at leaf nodes for classification. A combination of linear logistic regression and decision tree helps in dealing with the bias-variance tradeoff. This technique is robust to missing values and can handle numeric as well as nominal attributes.

- *Random Tree (RT)* (*Leo, 2001*; *Sumner, Frank & Hall, 2005*): Random Tree is an ensemble-based supervised learner where different trees are constructed from the same population. Random samples of the population are generated to form different trees with a random selection of features. After bags are constructed, models are developed and majority voting is performed to classify the class.
- *Random SubSpace method (RSS)* (*Ho, 1998*): Random SubSpace is used to construct random forests. Randomly feature subsets are selected to generate multiple trees. Bagging is performed with Reptree. Reptree is faster than the basic decision tree and generates multiple trees in each iteration. It then selects the tree whose performance is the best.

Decision trees

- *Pruning rule-based classification tree (PART)* (*Frank & Witten, 1998*): PART is a rule-based learning algorithm that exploits partial C4.5 decision trees and generates rules at each iteration. PART stands for a pruning rule-based classification tree. The rule that results in the best classification is selected. MDL is used to find the optimal split. smaller the confidence factor more will be the pruning done.
- *J48* (*Quinlan, 1993*): J48 is a JAVA implementation of the C4.5 decision tree. It follows the greedy technique to build a decision tree and uses the gain ratio as splitting criteria. Leaf nodes are the classification labels-defective and non-defective and rules can be derived by traversing from root to leaf node. It generates a binary tree, and one-third of the data is used for reduced error pruning.

Step 6: statistical validation

The results need to be statistically verified because without the involvement of statistics results may be misleading (*Frank & Witten, 1998*). *Kitchenham et al. (2017)* emphasized employing robust statistical tests for validating the experimental results. For statistical validation, we can use either parametric or nonparametric tests. The nonparametric Friedman statistical test (*Friedman, 1940*) and the Nemenyi Test are exercised in this study because software data do not follow a normal distribution (*Demšar, 2006*). The Friedman test is executed for different performance evaluators for establishing the statistical difference amongst the performance of developed SDP models. We need to compare several ML models built for several datasets. Therefore, Friedman rankings are computed using the Friedman test. Mean ranks are determined with the help of actual values of performance measure and then these ranks are exploited to perform post-hoc Nemenyi test. If the Friedman test results tend to be positive, post hoc analysis is carried by the Nemenyi test to find pair-wise significant differences. Nemenyi test is executed to determine the technique that statistically outperforms others. The Friedman test and the Nemenyi test are the non-parametric alternatives of the parametric ANOVA test and Tukey test. Both tests are performed using a 95% confidence interval. Hypotheses are set for the corresponding test and we need to accept or refute hypotheses at $\alpha = 0.05$.

Table 5 Parameters used in ML techniques.

Category	ML technique	Parameter settings
Statistical ML techniques	NB	useKernelEstimator = false, displayModelInOldFormat = false, useSupervisedDiscretization = false
	LR	Ridge: 1.0E−8, useConjugateGradientDescent = false, maxIts = −1
	SL	Heuristic Stop = 50, Max Boosting Iterations = 500, useCrossValidation = True, weightTrimBeta = 0.0
	LB	Zmax = 3.0, likelihoodThreshold = −1.7976931348623157E308, numIterations = 10, numThreads = 1, poolSize = 1, seed = 1, shrinkage = 1.0, useResampling = False, weightThreshold = 100
Neural Networks	MLP	Hidden layer = a, Learning rate = 0.3, Momentum = 0.2, Training time = 500, Validation threshold =20
Nearest Neighbour Methods	IBk	KNN = 1, nearestNeighbourSearchAlgorithm = LinearNNSearch,
	K*	globalBlend = 20, entropicAutoBlend = False
Ensemble Methods	ABM1	numIterations = 10, weightThreshold = 100, seed =1, classifier = J48: confidenceFactor = 0.25, minNumObj = 2, numFolds = 3, seed = 1, subtreeRaising = True, useMDLcorrection = True
	Bag	numIterations = 10, numExecutionSlots = 1, seed =1, bagSizePercent = 100, classifier = J48: confidenceFactor = 0.25, minNumObj = 2, numFolds = 3, seed = 1, subtreeRaising = True, useMDLcorrection = True
	ICO	evaluationMetric = RMSE, lookAheadIterations = 50, numFolds = 10, numRuns = 1, numThreads = 1, poolsize = 1, seed = 1, stepSize = 1, iterativeClassifier = LogitBoost
	LMT	errorOnProbabilities = False, fastRegression = True, minNumInstances = 15, numBoostingIterations = −1, weightTrimBeta = 0.0
	RT	KValue = 0, breakTiesRandomly = False, maxDepth = 0, minNum = 1, minVarianceProp = 0.001, numFolds = 0, seed = 1
	RSS	numExecutionSlots = 1, numIterations = 10, seed = 1, subSpaceSize = 0.5, classifier = Reptree: initialCount = 0.0, maxDepth = -1, minNum = 2.0, minVarianceProp = 0.001, numFolds = 3, seed = 1
Decision Tree	PART	confidenceFactor = 0.25, minNumObj = 2, numFolds = 3, reducedErrorPruning = False, seed = 1, useMDLcorrection = True
	J48	confidenceFactor = 0.25, minNumObj = 2, numFolds = 3, seed = 1, subtreeRaising = True, useMDLcorrection = True

RESULTS

RQ1: Which features are repeatedly selected by CFS in software engineering datasets?

A total of 20 OO metrics of datasets fundamentally define the IQAs of the software and can be grouped into cohesion, coupling, size, complexity, inheritance, encapsulation, and composition metrics. OO metrics corresponding to each IQA are presented in Table 6. #Selected denotes the number of times a particular metric is selected by CFS for all datasets. LCOM was selected by 11 datasets whereas the Ce metric was selected by 10 datasets. This RQ contemplates the metrics that are important for SDP. The weightage of each metric that was selected by CFS for 12 datasets is considered and their proportion selection was determined for each IQA.

In cohesion metrics, LCOM, CAM, and LCOM3 were chosen by 91.67%, 66.67%, and 41.67% of datasets respectively. The cumulative proportion of the selection of cohesion metrics is 66.7%. This shows that cohesion metrics are important for SDP and while developing the software, developers can focus more on LCOM and CAM values. Similarly, the composition metric (MOA) was selected by 66.67% of the datasets.

Table 6 Proportion selection of IQAs and CFS selected metrics.

IQA	OO Metric	#Selected	Proportion selection
Cohesion	LCOM	11	0.667
	CAM	8	
	LCOM3	5	
Composition	MOA	8	0.667
Size	WMC	6	0.583
	NPM	6	
	LOC	8	
	AMC	8	
Coupling	Ca	9	0.556
	Ce	10	
	CBO	5	
	RFC	9	
	CBM	3	
	IC	4	
Complexity	Avg_CC	4	0.5
	Max_CC	8	
Encapsulation	DAM	5	0.462
Inheritance	NOC	1	0.194
	MFA	3	
	DIT	3	

Exploring Table 6, though the cumulative proportion of coupling metrics is 55.6% its significance can be judged by selecting the top three selected metrics—RFC, Ca, and Ce. RFC is picked by 10 datasets whereas Ca and Ce are opted by 9 datasets each. Considering only these three metrics, the proportion selection of coupling metrics raises from 55.65% to 77.78%.

In size metrics, LOC and AMC are more preferred software metrics for defect prediction. In all datasets, the least selected metrics belong to the inheritance category. Therefore, resource investment can be done wisely by developers. The number of times any metric is selected for all the datasets guides developers and software practitioners in determining its worth for SDP.

RQ2: What is the performance of ML techniques on imbalanced data while building SDP models?

A box and whisker diagram (boxplot diagram) graphically represents numerical data distributions using five statistics: (a) the smallest observation, (b) lower quartile (Q_1), (c) median, (d) upper quartile (Q_3), and (e) the largest observation. The box is constructed based on the interquartile range (IQR) from Q_1 to Q_3. The median is represented by the line inside the box. The whiskers at both ends indicate the smallest observation and the largest observation.

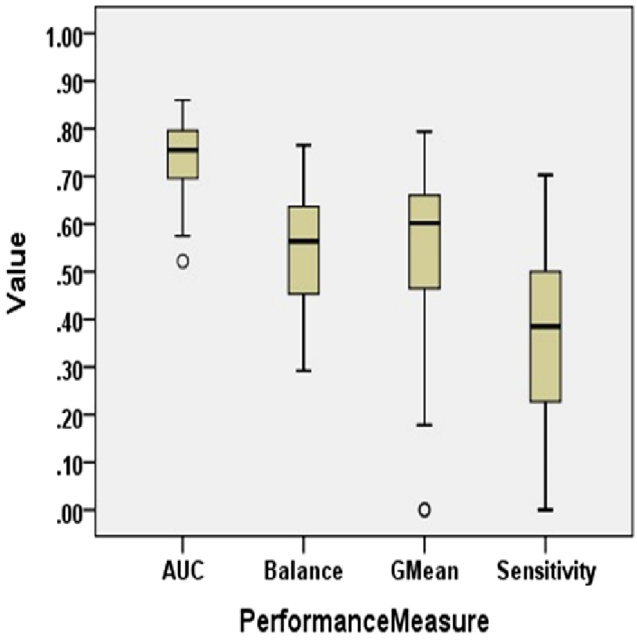


Figure 2 **Boxplot diagram representing performance measures using original data (NS Case).**
Full-size DOI: 10.7717/peerj-cs.573/fig-2

Figure 2 presents the Boxplot diagrams to visually depict the defect predictive capability of ML models on imbalanced data in terms of AUC, Balance, GMean, and Sensitivity.

AUC analysis

The AUC of ML models of imbalanced data varies from 0.52 to 0.86 with a median value of 0.76. Only 55.6% of models have AUC greater than 0.75. AUC is less than 0.85 in 97.8% of cases. Log4j1.1 depicted the best AUC value of 0.86 with NB. Analysis of results obtained on experimentation reveals that statistical techniques, NB and SL, have performed fairly well in terms of AUC, depicting the highest AUC values for Log4j1.1, Ant1.7, Jedit4.2, Log4j1.0, and Tomcat6.0. Similarly, LMT models were able to predict the highest AUC values for five datasets. Overall, 25% of ML models have AUC less than 0.7.

Balance analysis

The range of Balance in No Resampling (NS) case is from 29.20 to 76.55. In Tomcat6.0, only 59.01 value was achieved as the maximum value by NB. Balance values achieved by RSS, PART, and LMT are comparatively lower than other ML techniques. The median value for Balance for all models in the NS case is 56.35. Ivy2.0, Synape1.0, and Tomcat6.0 attained the maximum Balance value with NB. IBk also resulted in maximum Balance values for Jedit4.0, Synapse1.1, and Xerces1.3. 68.3% of datasets have a Balance value of less than 65. Only 3.3% of datasets have a Balance value greater than 75.

GMean analysis

When no resampling method is used, GMean ranges from 0 to 0.79. The median value of GMean for all datasets is observed as 0.6. NB achieved the highest GMean values for

58.33% of datasets. Considering the models for all datasets with 5 different ML techniques, only 14.4% of models achieved GMean greater than 0.7. 30.6% of models have a GMean value less than 0.5.

Sensitivity analysis

Sensitivity is less than 60% in 93.9% of cases. The median value of Sensitivity is only 0.39. This supports the low predictive capability of developed models when CIP is not handled. Sensitivity values lie between 0 and 0.63. Only 0.6% of cases have a Sensitivity greater than 70% which is not an acceptable achievement for any prediction model. The maximum Sensitivity value obtained in the NS case is 0.7 by IBk, the nearest neighbor technique, in the Log4j1.1 dataset.

Thus, the overall performance of SDP models developed using machine learning algorithms on imbalanced data is not satisfactory for high-quality predictions.

RQ3a: What is the comparative performance of various SDP models developed using resampling methods?

RQ3b: Is there any improvement in the performance of SDP models developed using ML techniques on the application of resampling methods?

To answer these questions, we exploited performance evaluators—Sensitivity, GMean, Balance, and AUC values that are calculated with help of a confusion matrix obtained by ten-fold cross-validation-trained models developed using resampling methods. Boxplot diagrams are generated and presented in Fig. 3 to visually depict the defect predictive capability of ML models in terms of AUC, Balance, GMean, and Sensitivity on resampled data. Median values of the best resampling method and NS scenario for AUC, Balance, GMean, and Sensitivity are recorded in Figs. 4, 5, 6, and 7 in form of bar graphs. We have analyzed the performance of developed models based on mean and median values obtained for the cumulative ML techniques of considered performance evaluators. NS cases are included to provide a fair comparison with the resampling-based models.

Comparative performance of various SDP models developed using resampling methods

For all the datasets, the median values are reported for NS and the resampling method that yields the maximum median performance value in bar graphs. Comparison of boxplot diagram in Fig. 2 with boxplot diagram in Fig. 3 shows the rise in the median line, quartiles, and the highest value achieved by models that were built using resampled data.

AUC analysis

On the application of resampling methods, 51.4% of models achieved AUC greater than 0.8. From the bars in Fig. 4, it is visible that ROS performance is the best amongst others. ROS attained the highest mean and median value for 75% of the datasets-Ant1.7, Camel1.6, Ivy2.0, Jedit4.0, Jedit4.2, Log4j1.0, Synapse1.0, Tomcat6.0, and Xerces1.3. It also showed the highest mean value for Synapse1.1. Other resampling methods like AHC, SMT,

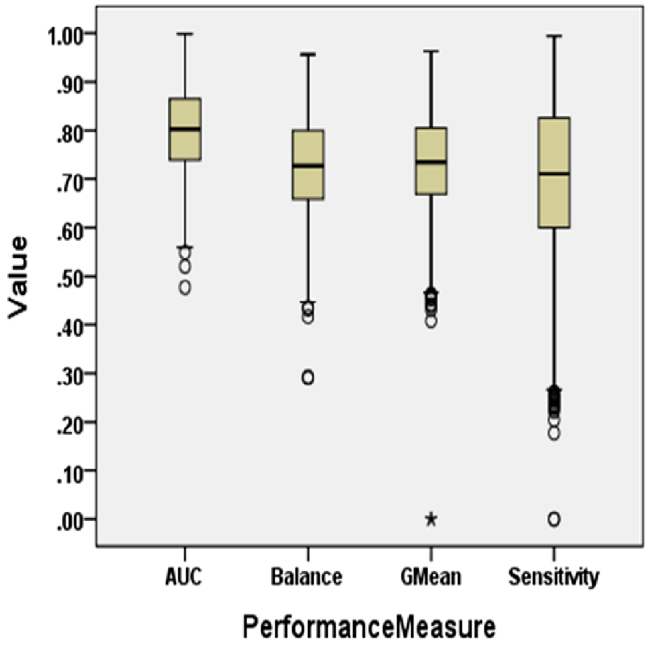


Figure 3 **Boxplot diagram representing performance measures using resampled data.**
Full-size DOI: 10.7717/peerj-cs.573/fig-3

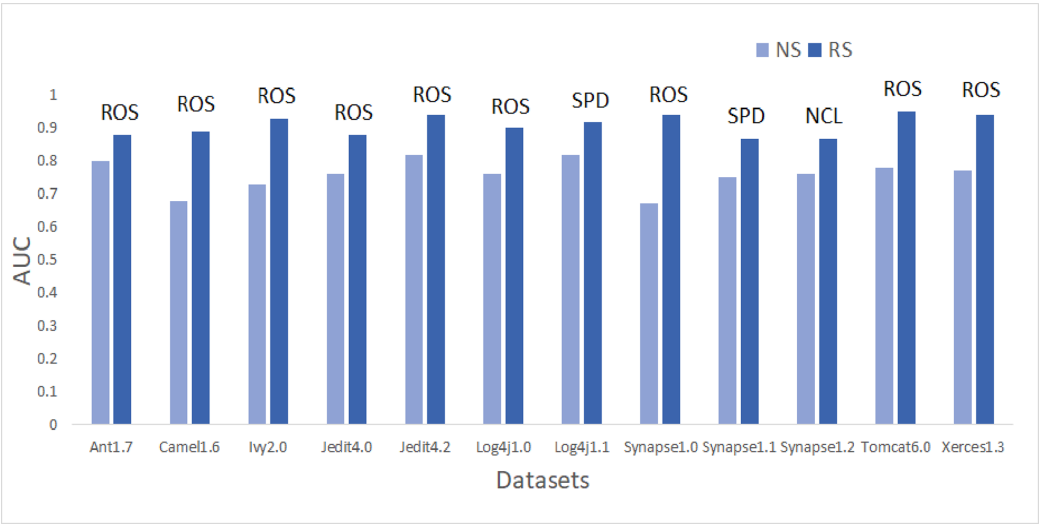


Figure 4 **Dataset-wise comparison of median AUC values of NS and RS.**
Full-size DOI: 10.7717/peerj-cs.573/fig-4

and SPD also have depicted good performance in terms of AUC. SPD got the highest median value for Log4j1.1 (0.92) and Synapse1.1 (0.87). In undersampling methods, only NCL results can be considered progressive. NCL was able to manage to secure the highest median value for only one dataset, i.e., Synapse1.2. The highest AUC value of 1 is

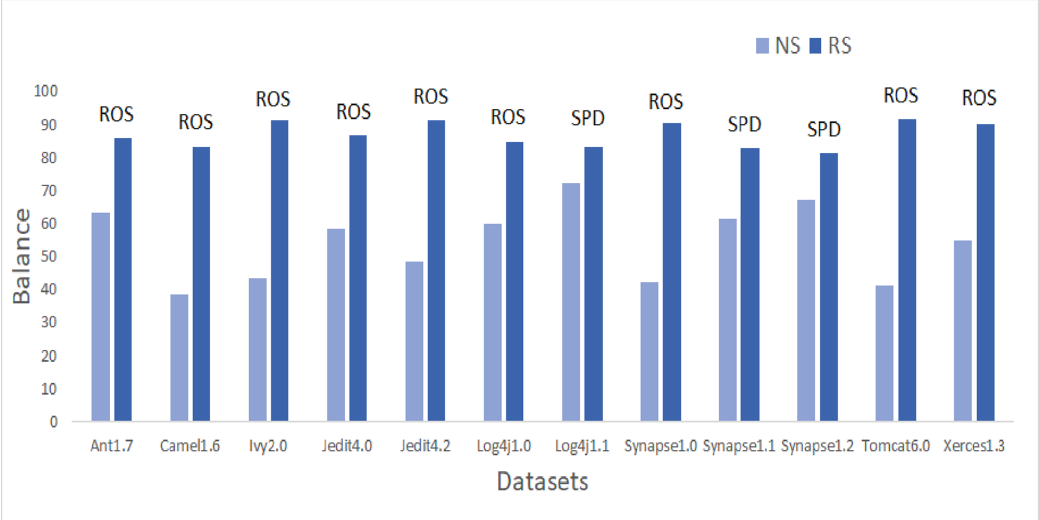


Figure 5 **Dataset-wise comparison of median balance values of NS and RS.**
Full-size DOI: 10.7717/peerj-cs.573/fig-5

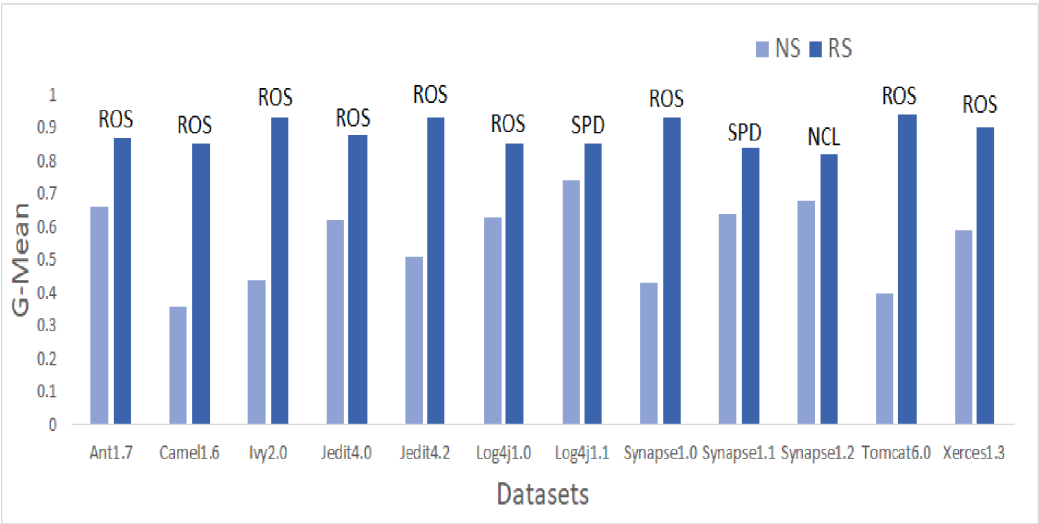


Figure 6 **Dataset-wise comparison of median GMean values of NS and RS.**
Full-size DOI: 10.7717/peerj-cs.573/fig-6

achieved by Jedit4.2 and Tomcat6.0 datasets. 13.9% of models have AUC greater than 90%, which is a remarkable improvement.

Balance analysis

Referring to Fig. 5, according to the Balance performance evaluator values achieved in predictive modeling, ROS performance seemed to outperform the other resampling methods. ADSYN could achieve a maximum of 86.24 Balance value for Synapse1.0. There were only three resampling methods, ROS, AHC, and SPD, that could achieve the highest

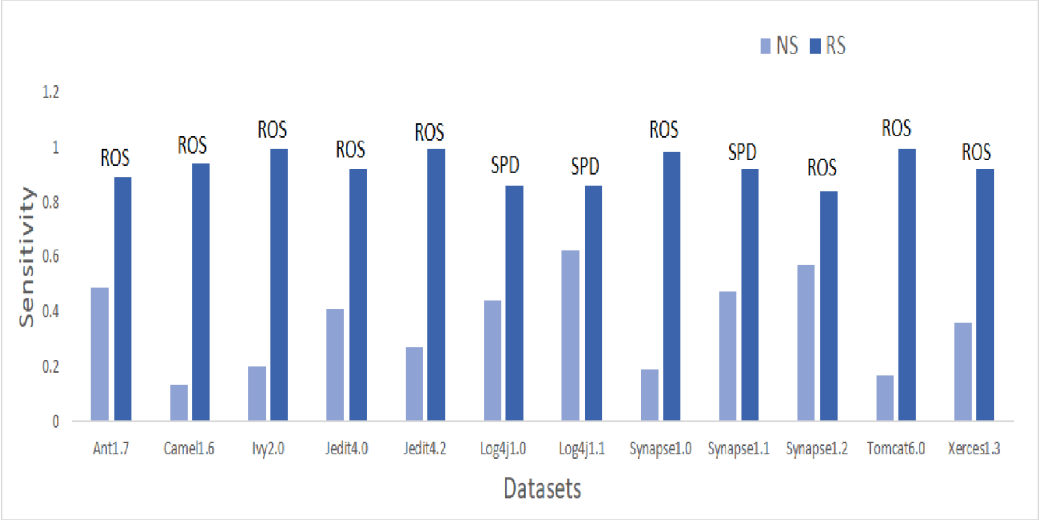


Figure 7 **Dataset-wise comparison of median sensitivity values of NS and RS.**
Full-size DOI: 10.7717/peerj-cs.573/fig-7

Balance value greater than 90. ROS was able to generate a Balance value of 95.58 for Ivy2.0. Comparatively, undersampling techniques, RUS and OSS, had a maximum Balance value of only 77.36 and 80.02 respectively. The highest median and mean values for all datasets except Synapse1.2 are attained by either SPD or ROS. Synapse 1.2 got the highest mean value with NCL and undersampling method. 60.9% of cases have a Balance greater than 70. Therefore, there was 357% of growth in median values of Balance when resampling is done as compared to NS.

GMean analysis

Similar patterns were also observed in GMean. Bar graphs in Fig. 6 portray the better predictive capabilities of ML techniques with ROS. The highest mean and median values are attained by ROS and SPD for all datasets except Synapse1.2. NCL gave the best results for mean and median values of Synapse1.2. 80.7% of cases have GMean greater than 0.65 for resampling methods. AHC, though not having any maximum values, but have consistent performance for all the datasets. With NS, only 3.9% of cases could achieve a GMean value greater than 0.75. With resampling methods, the number of cases for GMean values greater than or equal to 0.75 has elevated to 42.7% with the 1,182% of improvement. Comparing the oversampling and undersampling methods, oversampling methods thrived in predicting defects efficiently.

Sensitivity analysis

Sensitivity values are illustrated in the bar graph presented in Fig. 7. CNNTL worked best for Log4j1.1 and Synapse1.2 with an average Sensitivity value of 0.85 and 0.83 respectively. For other datasets, ROS outperformed other resampling methods. After applying resampling methods, Sensitivity increases above 0.9 for 11.9% of the developed models. 53% of resampling-based models have a Sensitivity greater than 0.7 as compared to only

Table 7 Percentage improvement in performance measures after resampling methods are used.

Datasets	AUC growth		Balance growth		GMean growth		Sensitivity growth	
	Max	Median	Max	Median	Max	Median	Max	Median
Ant1.7	18.1%	1.3%	30.9%	16.4%	27.8%	12.1%	67.2%	44.9%
Camel1.6	34.2%	4.4%	69.8%	61.3%	68.5%	72.2%	180.0%	353.8%
Ivy2.0	19.3%	13.7%	62.3%	74.7%	54.8%	75.0%	130.2%	250.0%
Jedit4.0	24.1%	1.3%	34.9%	23.6%	32.9%	17.7%	75.0%	73.2%
Jedit4.2	19.0%	1.2%	66.7%	49.5%	57.4%	45.1%	147.5%	155.6%
Log4j1.0	14.1%	9.2%	34.9%	26.0%	31.0%	20.6%	75.0%	68.2%
Log4j1.1	12.8%	0.0%	16.1%	2.2%	15.2%	1.4%	41.4%	19.4%
Synapse1.0	33.8%	23.9%	38.7%	86.8%	33.8%	86.0%	76.8%	326.3%
Synapse1.1	21.8%	0.0%	35.9%	13.1%	36.4%	9.4%	72.7%	44.7%
Synapse1.2	14.6%	3.9%	21.5%	7.2%	21.1%	5.9%	47.6%	33.3%
Tomcat6.0	23.5%	3.8%	70.8%	77.7%	62.7%	85.0%	160.5%	288.2%
Xerces1.3	16.9%	5.2%	45.8%	31.9%	38.8%	23.7%	93.9%	88.9%
Overall	16.3%	5.3%	24.9%	29.1%	21.5%	21.7%	41.4%	82.1%

0.6% of cases in the NS scenario. This accounts for a whopping 9,440% of improvement in ML models having a Sensitivity value greater than 0.7.

Therefore, it can be concluded that the predictive capability of ML models has immensely improved after treating imbalanced data properly using resampling methods.

Comparison of resampling-based ML models with NS models

A comparison of the results of models developed without employing resampling methods and models developed using resampling methods provides an answer to RQ3b. 1800 ML models were constructed for 12 datasets with ten resampling methods. The results were compared based on the maximum value attained and averaged median value achieved in each dataset. Percentage Improvement in Performance Measures after resampling methods are used is presented in Table 7. Analysis of Table 7 shows that there is a positive increment in all the values of maximum or median for the four performance evaluators when resampling methods are employed to build SDP models. This proves that there is a definite improvement in resampling-based ML models than the NS models when evaluated based on AUC, Balance, GMean, and Sensitivity.

For AUC, the overall percentage growth for the median value was 6.3%. The maximum AUC value achieved in the NS case is 0.86 for Log4j1.1 which increases to 0.97 when resampling methods are applied to it. Jedit4.2 and Tomcat6.0 were able to attain a maximum AUC value of 1 showing 18.2% and 22.2% of the increase. For Synapse1.0, on the application of resampling methods, the maximum value depicts the incremental growth of 33.8% and the respective median growth corresponds to 24.3%.

The increase in median values of Balance and GMean for all the datasets is 29.1% and 21.2% respectively with resampling methods. Synapse1.0, Tomcat6.0, Ivy2.0, and Camel1.6 has illustrated more than 60% of improvement in Balance median values and more than 70% improvement in GMean median values. The maximum Balance value

Table 8 Friedman rankings for resampling methods with AUC, Balance, GMean, and Sensitivity.

	Rank 1	Rank 2	Rank 3	Rank 4	Rank 5	Rank 6	Rank 7	Rank 8	Rank 9	Rank 10	Rank 11
AUC	**ROS (10.11)**	AHC (9.13)	SPD (8.58)	NCL (8.04)	SMT (7.06)	ADSYN (5.55)	NS (4.3)	SLSMT (4.05)	RUS (3.37)	CNNTL (3.08)	OSS (2.68)
Balance	**ROS (10.08)**	AHC (9.01)	SPD (7.93)	SMT (7.56)	NCL (6.6)	ADSYN (6.23)	SLSMT (5.09)	RUS (4.42)	CNNTL (3.85)	OSS (3.41)	NS (1.77)
GMean	**ROS (10.15)**	AHC (9.01)	SPD (7.91)	SMT (7.46)	NCL (6.95)	ADSYN (6.11)	SLSMT (4.88)	RUS (4.36)	CNNTL (3.66)	OSS (3.3)	NS (2.16)
Sensitivity	**ROS (9.89)**	AHC (8.83)	SMT (7.06)	SPD (7.06)	ADSYN (7.01)	CNNTL (6.48)	NCL (5.29)	SLSMT (5.2)	RUS (4.01)	OSS (3.9)	NS (1.22)

Note:
Best resampling technique is highlighted in bold.

gained by models with resampling methods is 95.58 for Ivy2.0 which was earlier 58.88. Similarly, the Sensitivity median values have shown a remarkable improvement of 84.6%. The median value of NS was 0.39 when all datasets were considered together. This value was raised to 0.71 on the application of resampling methods.

Answer to RQ3b—The results verify that there is an improvement in the performance of SDP models developed using ML techniques on the application of resampling methods.

RQ4: Which resampling method outperforms the addressed undersampling and oversampling techniques for building an efficient SDP model?

This RQ addresses the effectiveness of 10 resampling methods that are investigated in this study for constructing good SDP models. For the experiments conducted, is there any particular resampling method that can be considered the best? In this direction, we conducted Friedman tests on performance evaluators to provide rankings to models built using resampling methods. The case when no resampling method was used is also included.

The Friedman test was used to find the difference between techniques statistically. Four hypotheses are formed for four different performance evaluators. The hypothesis formed to achieve this objective is stated as:

H_i0 (Null Hypothesis): There is no significant statistical difference between the performance of any of the defect prediction models developed after using resampling methods and models developed using original data, in terms of PM_j.

$H_i a$ (Alternate Hypothesis): There is a significant statistical difference between the performance of any of the defect prediction models developed after using resampling methods and model developed using original data, in terms of PM_j.

where i = 1 to 4 denoting H_1, H_2, H_3 and H_4 hypothesis and j = 1 to 4. PM_1 = AUC, PM_2 = Balance, PM_3 = GMean, and PM_4 = Sensitivity.

Table 8 provides the desired ranking of SDP models developed in this study for AUC, Balance, GMean, and Sensitivity. Best resampling technique is highlighted in bold for each performance measure in Table 8. The mean ranks of each resampling method and NS are shown in parentheses. As discussed, NS represents the scenario when no sampling technique is used, so, it represents the performance with original data. We evaluated the hypothesis at the 0.05 level of significance, i.e., 95% of the confidence interval. Rank 1 is the best rank and rank 11 is the worst rank. The *p*-values achieved for all performance evaluators are 0.000 and recorded in Table 8. As *p*-values were less than 0.05, we rejected

Table 9 Pairwise comparison of resampling methods using Nemenyi test for AUC, Balance, GMean, and Sensitivity.

Pair	AUC	Balance	GMean	Sensitivity
ROS-AHC	NS	NS	NS	NS
ROS-SPD	S+	S+	S+	S+
ROS-SMT	S+	S+	S+	S+
ROS-NCL	S+	S+	S+	S+
ROS-ADSYN	S+	S+	S+	S+
ROS-SLSMT	S+	S+	S+	S+
ROS-RUS	S+	S+	S+	S+
ROS-CNNTL	S+	S+	S+	S+
ROS-OSS	S+	S+	S+	S+
ROS-NS	S+	S+	S+	S+

the null hypothesis and declared that there was a significant difference between resampling methods applied to developed SDP models.

Table 8 shows that ROS and AHC have unanimously scored Rank 1 and Rank 2 respectively for all the performance evaluators. Out of four undersampling methods, only one, i.e., NCL can make its space in the first seven positions for the reliable performance evaluators—AUC, Balance, and GMean. OSS, CNNTL, and RUS have acquired positions in the last four ranks with AUC, Balance, and GMean. These rankings clearly state the supremacy of oversampling methods over undersampling methods.

For Balance, GMean, and Sensitivity, NS case is ranked last. Therefore, results statistically approved the visualization in RQ3 that usage of resampling methods improved the predictive power of SDP models.

We have used Kendall's coefficient of concordance to assess the effect size. It is a quantitative measure of the magnitude of the experimental effect. Kendall in Table 8 shows the value for Kendall's coefficient of concordance. Its value ranges from 0 to 1. It reflects the degree of agreement. The Kendall value is 0.656, 0.589, 0.587, and 0.538 for AUC, Balance, GMean, and Sensitivity respectively. As the values are greater than 0.5 but less than 0.7, therefore, its effect is moderate.

The Friedman test shows whether there is an overall difference or not in the model performances, but if there is a difference, it fails to further identify the pairwise difference, i.e., exactly which technique is significantly different from other. For this, a post-hoc analysis was conducted using Nemenyi Test on overall datasets and the comparative pairwise performance of all the resampling methods was evaluated with ROS. The Nemenyi test was carried out at the $\alpha = 0.05$ level of significance. Table 9 summarizes the Nemenyi test results for AUC, GMean, Balance, and Sensitivity. 'S+' represents 'significantly better' results. If the difference between the mean ranks of the two techniques is less than the value of critical distance (CD), there is no significant difference in the 95% confidence interval. If the difference is greater than the CD value, the technique with a higher rank is considered statistically better. The computed CD value for the Nemenyi test conducted for resampling methods was 1.135. It can be inferred from Table 9 that ROS

Table 10 Friedman rankings for ML techniques with AUC, balance, gmean, and sensitivity.

	AUC	Balance	GMean	Sensitivity
Rank 1	**Kstar (14.41)**	**ABM1 (13.41)**	**ABM1 (13.5)**	**RT (12.62)**
Rank 2	ABM1 (13.75)	IBk (12.66)	IBk (12.91)	IBk (12.54)
Rank 3	BAG (13.37)	RT (12.16)	RT (12.58)	Kstar (12.33)
Rank 4	RSS (12.29)	BAG (12.12)	BAG (11.75)	ABM1 (12.2)
Rank 5	IBk (9.83)	Kstar (11.08)	Kstar (11.25)	LMT (10.95)
Rank 6	RT (9.2)	RSS (10.33)	RSS (9.58)	PART (9.95)
Rank 7	LMT (8.58)	PART (9.04)	PART (9.04)	J48 (9.83)
Rank 8	J48 (7.62)	LMT (9)	LMT (9)	BAG (9.54)
Rank 9	PART (7.58)	J48 (8.58)	J48 (8.7)	RSS (8.75)
Rank 10	LB (6.08)	MLP (5.45)	MLP (5.75)	LB (5.25)
Rank 11	ICO (5.54)	ICO (4.87)	ICO (4.87)	MLP (4.91)
Rank 12	MLP (4.75)	LB (4.83)	LB (4.7)	ICO (4.83)
Rank 13	LR (2.79)	SL (2.75)	SL (2.66)	SL (2.66)
Rank 14	SL (2.5)	LR (2.41)	LR (2.33)	LR (2.41)
Rank 15	NB (1.66)	NB (1.25)	NB (1.33)	NB (1.16)
p-value	**0.000**	**0.000**	**0.000**	**0.000**
Kendall	0.876	0.835	0.844	0.838

Note:
First ranked ML techniques and significant *p*-values attained for different performance measures are indicated in bold.

has comparable performance with AHC and exhibits statistically better performance than all other compared scenarios based on AUC, Balance, GMean, and Sensitivity.

Answer to RQ4: Oversampling methods resulted in better SDP models than undersampling methods. ROS and AHC emerged as the statistically better resampling method in terms of AUC, Balance, GMean, and Sensitivity.

RQ5: which ML technique performs the best for SDP based on resampled data?

To answer this RQ, we performed the Friedman test for 15 ML techniques by considering ROS-based models. The Friedman rankings are recorded in Table 10 with Rank 1 as the best rank. The ML Technique that achieved best rank is bold in Table 10. The Friedman test was held at the 0.05 significance level with a degree of freedom of 14. The null hypothesis is set as there is no difference between comparative performances of ROS-based models for different ML techniques. In Table 10, significant *p*-values (p-values < 0.05) are highlighted in bold. The *p*-value for each performance evaluator was 0.000. Therefore, the results are considered 95% significant. We refute the null hypothesis as there is a significant difference amongst performances of models built using different ML techniques. The Kendall value for AUC, Balance, GMean, and Sensitivity for ML techniques with ROS-based models is 0.876, 0.835, 0.844, and 0.838. All the four performance evaluators have a Kendall value greater than or equal to 0.835. This signifies that rankings for different datasets are approximate 83.5% similar and hence, increases the

Table 11 Pairwise comparison of ML techniques using Nemenyi Test for AUC, Balance, GMean, and Sensitivity.

Pair	AUC	Balance	GMean	Sensitivity
ABM1-IBk	NS	NS	NS	NS
ABM1-RT	NS	NS	NS	NS
ABM1-BAG	NS	NS	NS	NS
ABM1-Kstar	NS	NS	NS	NS
ABM1-RSS	NS	NS	NS	NS
ABM1-PART	NS	NS	NS	NS
ABM1-LMT	NS	NS	NS	NS
ABM1-J48	NS	NS	NS	NS
ABM1-MLP	**S+**	**S+**	**S+**	**S+**
ABM1-ICO	**S+**	**S+**	**S+**	**S+**
ABM1-LB	**S+**	**S+**	**S+**	**S+**
ABM1-SL	**S+**	**S+**	**S+**	**S+**
ABM1-LR	**S+**	**S+**	**S+**	**S+**
ABM1-NB	**S+**	**S+**	**S+**	**S+**

Note:
Significant results are in bold.

reliability and credibility of the Friedman statistical results. The impact of differences in results is high.

This can be observed from Table 10 that Kstar, IBk, ABM1, RT, and RSS techniques incited better prediction models than other ML techniques.

Kstar and IBk are the variants of the nearest neighbor techniques. These ML techniques provide good results when there is a large number of samples irrespective of the data distribution. Nearest neighbors are instance-based fast learners. Mean ranks are inscribed in brackets. The mean rank of KStar is 14.41 for AUC which is the highest in the pool. The rank of IBk is second when the models are evaluated based on Balance, GMean, and Sensitivity. The mean rank of IBk is 12.66, 12.91, and 12.33 for Balance, GMean, and Sensitivity respectively.

ABM1, RT, Bag, and RSS have also attained high ranks. ABM1 is the first ranker in Balance and GMean with a mean rank of 13.41 and 13.5 respectively. ABM1 got 2nd rank with AUC and 4th rank with sensitivity. For the stable performance evaluators, ABM1, Bag, RT, and RSS appeared in the first six ranks, proving their competency in the ML world. These four techniques are ensembles that are considered robust in dealing with imbalanced data. The crux of ensemble techniques is to cover the weakness of the base ML technique and combine them to reduce bias or variance and enhance its predictive capability.

The last three ranks for all the performance evaluators are grabbed by the statistical learners: Naïve Bayes, Simple Logistic, and Logistic Regression. These ML techniques, in contrast, were able to generate good prediction models when no resampling method was involved in model construction. The balancing of both the classes boosted the performance of other classifiers, especially ensembles and nearest neighbors, and resulted in their unacceptable performance.

The post-hoc analysis is carried out using the Nemenyi Test and results are computed with CD = 6.191. Pairwise Comparison of ML Techniques using Nemenyi Test for AUC, Balance, GMean, and Sensitivity are presented in Table 11 and significant results are in bold. ABM1 was paired with other ML techniques and it was found comparable with IBk, Kstar, RT, BAG, RSS, PART, LMT, and J48. ABM1 was found statistically significant than MLP, ICO, LB, SL, LR, and NB.

Answer to RQ5: Ensembles and nearest neighbors performed the best for SDP in imbalanced data with a random oversampling method.

DISCUSSIONS

Feature Selection is an important activity in software quality predictive modeling. The most widely accepted feature selection technique in literature is CFS. Through this study, we provided the important subset of features to the software practitioners. We identified that coupling and cohesion are the most important internal quality attributes that developers should focus on while building software to avoid defects. Defects in software were found to be highly correlated to LCOM, RFC, Ca, and Ce.

Median values of Sensitivity have improved by 45.7%, 341.9%, 250%, 72.2%, 154.8%, 66.7%, 18.9%, 330%, 46.5%, 32.8%,293.3%, and 88.8% for Ant1.7, Camel1.6, Ivy2.0, Jedit4.0, Jedit4.2, Log4j1.0, Log4j1.1, Synapse1.0, Synapse1.1, Synapse1.2, Tomcat6.0, and Xerces1.3 respectively after applying resampling methods.

Median values of Balance have improved by 16.4%, 61.3%, 74.7%, 23.6%, 49.5%, 26%, 2.2%, 86.8%, 13.1%, 7.2%, 77.7%, and 31.9% for Ant1.7, Camel1.6, Ivy2.0, Jedit4.0, Jedit4.2, Log4j1.0, Log4j1.1, Synapse1.0, Synapse1.1, Synapse1.2, Tomcat6.0, and Xerces1.3 respectively after applying resampling methods.

Median values of G-Mean have improved by 13.1%, 73.7%, 73.5%, 16.9%, 44.7%, 20.1%, 0.4%, 88.3%, 9%, 5.5%, 83.5%, and 24.2% for Ant1.7, Camel1.6, Ivy2.0, Jedit4.0, Jedit4.2, Log4j1.0, Log4j1.1, Synapse1.0, Synapse1.1, Synapse1.2, Tomcat6.0, and Xerces1.3 respectively after applying resampling methods.

Median values of ROC-AUC have improved by 1.9%, 3.8%, 14.3.7%, 1.9%, 1.3%, 8.7%, 0.6%, 24.3%, 0.7%, 3.3%, 4.7%, and 5.8% for Ant1.7, Camel1.6, Ivy2.0, Jedit4.0, Jedit4.2, Log4j1.0, Log4j1.1, Synapse1.0, Synapse1.1, Synapse1.2, Tomcat6.0, and Xerces1.3 respectively after applying resampling methods.

For the datasets considered, ROS and AHC were significantly better than other resampling methods. Better model predictions can be achieved by incorporating oversampling methods than the undersampling methods. Our findings are in the agreement with the *Wang & Yao (2013)* conclusions. They proved that resampling methods improve defect prediction and, in their settings, the AdaBoost ensemble gave the best performance.

Resampling methods did not improve the defect predictive capability of statistical techniques. Developers and researchers should prefer ensemble methods for software quality predictive modeling.

VALIDITY THREATS

Conclusion validity

Conclusion validity threat is a threat to statistical validity and this indicates that results of the empirical study are not properly analyzed and validated. To avoid this threat, ten-fold cross-validation is performed. The Friedman test and Nemenyi test during post hoc analysis strengthens conclusion validity further. In this study, both the statistical validation techniques are nom parametric in nature. Non-parametric tests are not based on any assumptions for underlying data and, therefore, applicable to the selected datasets. This reinforces the analysis of the relationship between independent variables and dependent variables and hence enhancing conclusion validity.

Internal validity

Applying resampling methods results in the change in the original ratio of defective and non-defective classes. This would be affecting the causal relationship between independent and dependent variables resulting in internal validity bias in our study. However, we have used stable performance evaluators, GMean and AUC, to assess the performance of different SDP that help to reduce this threat. We have also used Sensitivity that takes care of the proper classification of defective classes, which is one of the major requirements of our problem domain. So, judicious selection of performance evaluators may have reduced some effect of internal validation threat.

Construct validity

Construct validity ensures the correctness of the way of measuring the independent and dependent variables of the study. It also emphasizes whether the variables are correctly mapped to the concept that they are representing. The Independent variables of this study are object-oriented software metrics and the dependent variable of this study is 'Defect' representing the absence or presence of a defect in the software module. Chosen independent variables and dependent variables are widely used in defect prediction studies and thus builds confidence in the removal of construct validity threats from our study. Hence, any construct validity threats may not exist in our study.

External validity

External validity refers to the extent to which the results of the study are widely applicable. Whether the conclusions of the concerned study can be generalized or not? The datasets used in this study are available publicly. Concerned software are written in the JAVA language. Thus, results validity holds for similar situations only. Results may not be valid for proprietary software. Resampling methods are implemented with default parameter settings in KEEL (*Alcalá-Fdez et al., 2011*), and ML parameter settings are provided. Therefore, this study can be reproduced without any complications. This minimizes the external validity threat in this study.

CONCLUSIONS

This study evaluates the effect of resampling methods on various machine learning models for defect prediction using Apache software. In total, 1,980 models were built and the performances of models were empirically compared using stable performance evaluators such as GMean, Balance, and AUC. Important features were selected using CFS and ten-fold cross-validation was executed while training the model. Six oversampling and four undersampling methods were analyzed with 15 ML techniques and results were statistically validated by using the Friedman test followed by Nemenyi *post hoc* analysis. The use of statistical tests reinforces the correctness of results. CFS is incorporated to minimize the multicollinearity effect whereas resampling methods reduces the model bias and assures that predictions are not affected by majority class label. This study reinsures that SDP models developed with resampling methods enhance their predictive capability as compared to models developed without resampling methods. The results of the Friedman test strongly advocate the use of the random oversampling method for the improved predictive capability of SDP classifiers for imbalanced data. Apart from ROS, AHC and SMT also demonstrated good predictive capability for uncovering defects. The Nemenyi test eradicates family-wise error and concluded ROS and AHC to be significantly and statistically better than other resampling methods. Models developed using oversampling methods illustrated better defect prediction capability than that of undersampling methods. Handling CIP using ROS and AHC will aid developers and software practitioners in detecting defects effectively in the early stages of software development reducing testing cost and effort. Resampling methods greatly improved the performance of ensemble methods.

Future direction involves the inclusion of more imbalanced datasets and investigates the impact of resampling methods on them. Datasets can be taken of any prevalent language different than JAVA like C#. There is a dire need for a benchmark study that compares all the existing resampling solutions in the literature with the common experimental settings. Optimized versions of base resampling methods like RUS or ROS can be proposed. Further, we will also like to explore the consequences of resampling methods with search-based techniques for the classification of software defects. Instead of predicting defects, a framework can besides be utilized to envisage defect severity.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

The authors received no funding for this work.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

- Ruchika Malhotra conceived and designed the experiments, analyzed the data, performed the computation work, authored or reviewed drafts of the paper, supervision, and approved the final draft.

- Juhi Jain conceived and designed the experiments, performed the experiments, analyzed the data, performed the computation work, prepared figures and/or tables, authored or reviewed drafts of the paper, and approved the final draft.

Data Availability

The following information was supplied regarding data availability:

Raw data are available as Supplemental Files 1 to 12. These files represent the datasets used in this study.

Regarding code availability: We used KEEL and WEKA software for implementing resampling methods and ML techniques. Both these tools are opensource tools.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj-cs.573#supplemental-information.

REFERENCES

Agrawal A, Menzies T. 2018. Is "Better Data" better than "Better Data Miners"? In: *IEEE/ACM 40th International Conference on Software Engineering*. 1050–1061.

Aha DW, Kibler D, Albert MK. 1991. Instance-based learning algorithms. *Machine Learning* **6(1)**:37–66.

Alcalá-Fdez J, Fernández A, Luengo J, Derrac J, García S, Sánchez L, Herrera F. 2011. Keel data-mining software tool: data set repository, integration of algorithms and experimental analysis framework. *Journal of Multiple-Valued Logic & Soft Computing* **1(6)**:17.

Arar ÖF, Ayan K. 2017. A feature dependent naive Bayes approach and its application to the software defect prediction problem. *Applied Soft Computing* **59(10)**:197–209 DOI 10.1016/j.asoc.2017.05.043.

Balogun AO, Basri S, Abdulkadir SJ, Hashim AS. 2019. Performance analysis of feature selection methods in software defect prediction: a search method approach. *Applied Sciences* **9(13)**:2764 DOI 10.3390/app9132764.

Balogun AO, Basri S, Mahamad S, Abdulkadir SJ, Almomani MA, Adeyemo VE, Al-Tashi Q, Mojeed HA, Imam AA, Bajeh AO. 2020. Impact of feature selection methods on the predictive performance of software defect prediction models: an extensive empirical study. *Symmetry* **12(7)**:1147 DOI 10.3390/sym12071147.

Bansiya J, Davis CG. 2002. A hierarchical model for object-oriented design quality assessment. *IEEE Transactions on Software Engineering* **28(1)**:4–17 DOI 10.1109/32.979986.

Basili VR, Briand LC, Melo WL. 1996. A validation of object-oriented design metrics as quality indicators. *IEEE Transactions on Software Engineering* **22(10)**:751–761 DOI 10.1109/32.544352.

Batista GE, Prati RC, Monard MC. 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explorations Newsletter* **6(1)**:20–29 DOI 10.1145/1007730.1007735.

Bennin KE, Keung J, Monden A, Kamei Y, Ubayashi N. 2016. Investigating the effects of balanced training and testing datasets on effort-aware fault prediction models. In: *IEEE 40th Annual Computer Software and Applications Conference.* Vol. 1. 154–163.

Bennin KE, Keung J, Phannachitta P, Monden A, Mensah S. 2018. Mahakil: Diversity based oversampling approach to alleviate the class imbalance issue in software defect prediction. *IEEE Transactions on Software Engineering* **44(6)**:534–550 DOI 10.1109/TSE.2017.2731766.

Breiman L. 1996. Bagging predictors. *Machine Learning* **24(2)**:123–140.

Bunkhumpornpat C, Sinapiromsaran K, Lursinsap C. 2009. Safe-level-smote: safe-level-synthetic minority over-sampling technique for handling the class imbalanced problem. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Vol. 27. 475–482.

Burez J, Van den Poel D. 2009. Handling class imbalance in customer churn prediction. *Expert Systems with Applications* **36(3)**:4626–4636 DOI 10.1016/j.eswa.2008.05.027.

Chawla NV, Bowyer KW, Hall LO, Kegelmeyer WP. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research* **16(6)**:321–357 DOI 10.1613/jair.953.

Chidamber SR, Kemerer CF. 1994. A metrics suite for object oriented design. *IEEE Transactions on Software Engineering* **20(6)**:476–493 DOI 10.1109/32.295895.

Cleary JG, Trigg LE. 1995. K*: an instance-based learner using an entropic distance measure. In: *Machine Learning Proceedings*. 108–114.

Cohen G, Hilario M, Sax H, Hugonnet S, Geissbuhler A. 2006. Learning from imbalanced data in surveillance of nosocomial infection. *Artificial Intelligence in Medicine* **37(1)**:7–18 DOI 10.1016/j.artmed.2005.03.002.

Demšar J. 2006. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine Learning Research* **7(1)**:1–30.

Fawcett T. 2006. An introduction to ROC analysis. *Pattern Recognition Letters* **27(8)**:861–874 DOI 10.1016/j.patrec.2005.10.010.

Frank E, Witten IH. 1998. Generating accurate rule sets without global optimization. In: *Fifteenth International Conference on Machine Learning*. 144–151.

Freund Y, Schapire RE. 1996. Experiments with a new boosting algorithm. In: *13th International Conference on Machine Learning*. Vol. 96. 148–156.

Friedman J, Hastie T, Tibshirani R. 2000. Additive logistic regression: a statistical view of boosting (with discussion and a rejoinder by the authors). *The Annals of Statistics* **28(2)**:337–407 DOI 10.1214/aos/1016218223.

Friedman M. 1940. A comparison of alternative tests of significance for the problem of m rankings. *The Annals of Mathematical Statistics* **11(1)**:86–92 DOI 10.1214/aoms/1177731944.

Galar M, Fernandez A, Barrenechea E, Bustince H, Herrera F. 2011. A review on ensembles for the class imbalance problem: bagging-, boosting-, and hybrid-based approaches. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* **42(4)**:463–484 DOI 10.1109/TSMCC.2011.2161285.

Ghotra B, McIntosh S, Hassan AE. 2017. A large-scale study of the impact of feature selection techniques on defect classification models. In: *IEEE/ACM 14th International Conference on Mining Software Repositories (MSR)*. Vol. 20. Piscataway: IEEE, 146–157.

He H, Bai Y, Garcia EA, Li S. 2008. ADASYN: adaptive synthetic sampling approach for imbalanced learning. In: *IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*. Vol. 1. Piscataway: IEEE, 1322–1328.

Henderson-Sellers B. 1995. *Object-oriented metrics: measures of complexity*. Hoboken: Prentice-Hall.

Ho TK. 1998. The random subspace method for constructing decision forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **20(8)**:832–844 DOI 10.1109/34.709601.

Jindaluang W, Chouvatut V, Kantabutra S. 2014. Under-sampling by algorithm with performance guaranteed for class-imbalance problem. In: *International Computer Science and Engineering Conference*. Vol. 30. 215–221.

John GH, Langley P. 1995. Estimating continuous distributions in Bayesian classifiers. In: *Proceedings of the Eleventh Conference on Uncertainty in Artificial Intelligence*. 338–345.

Jureczko M, Madeyski L. 2010. Towards identifying software project clusters with regard to defect prediction. In: *Proceedings of the 6th International Conference on Predictive Models in Software Engineering*. 1–10.

Jureczko M, Spinellis D. 2010. Using Object-Oriented Design metrics to predict software defects. In: *Models and Methods of System Dependability*. Wrocław: Oficyna Wydawnicza Politechniki Wrocławskiej, 69–81.

Kamei Y, Monden A, Matsumoto S, Kakimoto T, Matsumoto KI. 2007. The effects of over and under sampling on fault-prone module detection. In: *First International Symposium on Empirical Software Engineering and Measurement*. 196–204.

Khalilia M, Chakraborty S, Popescu M. 2011. Predicting disease risks from highly imbalanced data using random forest. *BMC Medical Informatics and Decision Making* **11(1)**:51 DOI 10.1186/1472-6947-11-51.

Khoshgoftaar TM, Gao K. 2009. Feature selection with imbalanced data for software defect prediction. In: *International Conference on Machine Learning and Applications*. Vol. 13. 235–240.

Khoshgoftaar TM, Gao K, Seliya N. 2010. Attribute selection and imbalanced data: problems in software defect prediction. In: *22nd IEEE International Conference on Tools with Artificial Intelligence*. Vol. 1. 137–144.

Kitchenham B, Madeyski L, Budgen D, Keung J, Brereton P, Charters S, Gibbs S, Pohthong A. 2017. Robust statistical methods for empirical software engineering. *Empirical Software Engineering* **22(2)**:579–630 DOI 10.1007/s10664-016-9437-5.

Koru AG, Tian J. 2005. Comparing high-change modules and modules with the highest measurement values in two large-scale open-source products. *IEEE Transactions on Software Engineering* **31(8)**:625–642 DOI 10.1109/TSE.2005.89.

Kotsiantis S, Kanellopoulos D, Pintelas P. 2006. Handling imbalanced datasets: a review. *GESTS International Transactions on Computer Science and Engineering* **30(1)**:25–36.

Kubat M, Matwin S. 1997. Addressing the curse of imbalanced training sets: one-sided selection. In: *International Conference of Machine Learning*. Vol. 97. 179–186.

Landwehr N, Hall M, Frank E. 2005. Logistic model trees. *Machine learning* **59(1–2)**:161–205 DOI 10.1007/s10994-005-0466-3.

Lane PC, Clarke D, Hender P. 2012. On developing robust models for favourability analysis: model choice, feature sets and imbalanced data. *Decision Support Systems* **53(4)**:712–718 DOI 10.1016/j.dss.2012.05.028.

Laurikkala J. 2001. Improving identification of difficult small classes by balancing class distribution. In: *Conference on Artificial Intelligence in Medicine in Europe*. 63–66.

Le Cessie S, Van Houwelingen JC. 1992. Ridge estimators in logistic regression. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* **41(1)**:191–201.

Leo B. 2001. Random forests. *Machine Learning* **45(1)**:5–32 DOI 10.1023/A:1010933404324.

Lessmann S, Baesens B, Mues C, Pietsch S. 2008. Benchmarking classification models for software defect prediction: a proposed framework and novel findings. *IEEE Transactions on Software Engineering* **34(4)**:485–496 DOI 10.1109/TSE.2008.35.

Li M, Zhang H, Wu R, Zhou ZH. 2012. Sample-based software defect prediction with active and semi-supervised learning. *Automated Software Engineering* **19(2)**:201–230 DOI 10.1007/s10515-011-0092-1.

Lingden P, Alsadoon A, Prasad PW, Alsadoon OH, Ali RS, Nguyen VT. 2019. A novel modified undersampling (MUS) technique for software defect prediction. *Computational Intelligence* **35(4)**:1003–1020 DOI 10.1111/coin.12229.

Liu Y, An A, Huang X. 2006. Boosting prediction accuracy on imbalanced datasets with SVM ensembles. In: *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Vol. 9. 107–118.

Malhotra R. 2015. A systematic review of machine learning techniques for software fault prediction. *Applied Soft Computing* **27**:504–518 DOI 10.1016/j.asoc.2014.11.023.

Malhotra R, Kamal S. 2019. An empirical study to investigate oversampling methods for improving software defect prediction using imbalanced data. *Neurocomputing* **343(2)**:120–140 DOI 10.1016/j.neucom.2018.04.090.

Malhotra R, Khanna M. 2017. An empirical study for software change prediction using imbalanced data. *Empirical Software Engineering* **22(6)**:2806–2851 DOI 10.1007/s10664-016-9488-7.

Malhotra R. 2016. *Empirical research in software engineering: concepts, analysis, and applications*. New York: Chapman and Hall/CRC.

Martin R. 1994. OO design quality metrics. *An Analysis of Dependencies* **12(1)**:151–170.

Mazurowski MA, Habas PA, Zurada JM, Lo JY, Baker JA, Tourassi GD. 2008. Training neural network classifiers for medical decision making: the effects of imbalanced datasets on classification performance. *Neural Networks* **21(2–3)**:427–436 DOI 10.1016/j.neunet.2007.12.031.

Moreo A, Esuli A, Sebastiani F. 2016. Distributional random oversampling for imbalanced text classification. In: *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*. Vol. 7. New York: ACM, 805–808.

Pelayo L, Dick S. 2007. Applying novel resampling strategies to software defect prediction. *NAFIPS Annual Meeting of the North American Fuzzy Information Processing Society* **24(6)**:69–72.

Pelayo L, Dick S. 2012. Evaluating stratification alternatives to improve software defect prediction. *IEEE Transactions on Reliability* **61(2)**:516–525 DOI 10.1109/TR.2012.2183912.

Phua C, Alahakoon D, Lee V. 2004. Minority report in fraud detection: classification of skewed data. *ACM SIGKDD Explorations Newsletter* **6(1)**:50–59 DOI 10.1145/1007730.1007738.

Quinlan JR. 1993. *C4.5: programs for machine learning*. San Mateo: Morgan Kaufmann.

Radjenović D, Heričko M, Torkar R, Živkovič A. 2013. Software fault prediction metrics: a systematic literature review. *Information and Software Technology* **55(8)**:1397–1418 DOI 10.1016/j.infsof.2013.02.009.

Riquelme JC, Ruiz R, Rodríguez D, Moreno J. 2008. Finding defective modules from highly unbalanced datasets. *Actas de los Talleres de las Jornadas de Ingeniería del Software y Bases de Datos* **2(1)**:67–74.

Rodriguez D, Herraiz I, Harrison R, Dolado J, Riquelme JC. 2014. Preliminary comparison of techniques for dealing with imbalance in software defect prediction. In: *Proceedings of the 18th International Conference on Evaluation and Assessment in Software Engineering*. Vol. 13. 1–10.

Rojas R, Feldman J. 2013. *Neural networks: a systematic introduction*. Berlin: Springer-Verlag.

Seiffert C, Khoshgoftaar TM, Van Hulse J, Folleco A. 2014. An empirical study of the classification performance of learners on imbalanced and noisy software quality data. *Information Sciences* **259(2)**:571–595 DOI 10.1016/j.ins.2010.12.016.

Shatnawi R. 2012. Improving software fault-prediction for imbalanced data. In: *International Conference on Innovations in Information Technology*. Vol. 18. 54–59.

Singh Y, Kaur A, Malhotra R. 2010. Empirical validation of object-oriented metrics for predicting fault proneness models. *Software Quality Journal* **18(1)**:3–35 DOI 10.1007/s11219-009-9079-6.

Stefanowski J, Wilk S. 2008. Selective pre-processing of imbalanced data for improving classification performance. In: *International Conference on Data Warehousing and Knowledge Discovery*. Vol. 2. 283–292.

Sumner M, Frank E, Hall M. 2005. Speeding up logistic model tree induction. In: *European Conference on Principles of Data Mining and Knowledge Discovery*. 675–683.

Tan M, Tan L, Dara S, Mayeux C. 2015. Online defect prediction for imbalanced data. In: *IEEE/ACM 37th IEEE International Conference on Software Engineering*. Vol. 2. Prentice-Hall, 99–108.

Tantithamthavorn C, Hassan AE, Matsumoto K. 2018. The impact of class rebalancing techniques on the performance and interpretation of defect prediction models. *IEEE Transactions on Software Engineering* **46(11)**:1200–1219 DOI 10.1109/TSE.2018.2876537.

Tomek I. 1976. Two modifications of CNN. *IEEE Transactions on Systems, Man, and Cybernetics: Systems* **6**:769–772.

Vasu M, Ravi V. 2011. A hybrid under-sampling approach for mining unbalanced datasets: applications to banking and insurance. *International Journal of Data Mining, Modelling and Management* **3(1)**:75–105 DOI 10.1504/IJDMMM.2011.038812.

Wang S, Yao X. 2013. Using class imbalance learning for software defect prediction. *IEEE Transactions on Reliability* **62(2)**:434–443 DOI 10.1109/TR.2013.2259203.

Zheng Z, Wu X, Srihari R. 2004. Feature selection for text categorization on imbalanced data. *ACM SIGKDD Explorations Newsletter* **6(1)**:80–89 DOI 10.1145/1007730.1007741.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

PREDICTION OF CUTTING ENERGY AND FORCES ENCOUNTERED IN TURNING PROCESS USING CAD SOFTWARE

A.K. Madan*1, Aditya Kumar*2, Yash Kumar*3

*1Professor, Department Of Mechanical Engineering, Delhi Technological University, Delhi, India.

*2,3Student, Department Of Mechanical Engineering, Delhi Technological University, Delhi, India.

ABSTRACT

In this research paper the cutting forces and torque produced during the turning operation are analyzed including the effect of the chattering and damping in the cutting process with the use of Computer aided Manufacturing. It uses different software to control machine tools in the manufacturing of work pieces. Cutting forces are crucial in cutting operations as they affect cutting performance such as surface finish, tool wear, cutting temperature, tool breakage, chattering etc. The tool geometry is very important in determining cutting forces etc. For example, the most important angle is the rake angle. If this angle is not made properly, it will affect its tool life as well as surface finish of the workpiece and various other aspects as well. In this paper, the center focus will revolve around the different implications of different rake angles on Cutting forces, Power requirement and the effects of cutting speed and feed rate upon thrust force.

Keywords: Tool, Turning, Rake Angle, CAM, Manufacturing, Cutting Forces, Feed Rate.

I. INTRODUCTION

Modelling and simulation are among the interesting ways of assessing and exploring metal cutting processes. Despite of several advancements in machine tool architecture, monitoring, and control process many scientists and industrialists are still working on optimizing the machining process in various research facilities[2]. For example, various models are designed or fabricated to predict the cutting forces, residual stresses, hardness, tool-wear, temperatures, part distortion, process dynamics, surface roughness, chip breaking/breakability, machining operations stability, which helps in judgement of overall machining performances.

Cutting forces have a complicated spatial placement inside the coordinate system associated to the tool's deformation axes. The following force decomposition for [X, Y, Z] is used to examine forces created during cutting: This relationships between cutting force components will be valid for scenario shown in fig (see Figure 1).[5]

$$R = \sqrt{P_x^2 + P_y^2 + P_z^2}$$

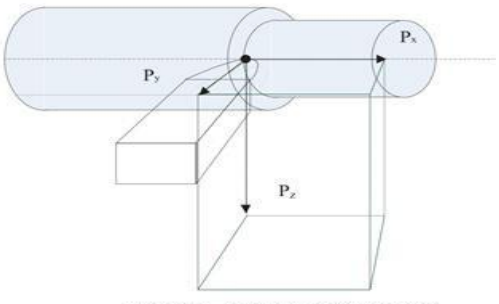


FIGURE 1. Generation of Cutting Forces

P represents the projections of the cutting force on X, Y, and Z as P_X, P_Y, P_Z and R is the net force (applied against the shape-generating movements). The relationship between the components in question is influenced by a variety of parameters, including geometry, edge wear, and so on[5]. Component Pz, also known as the tangential cutting force, has the highest value.

RAKE ANGLE

The rake angle is the angle between the rake surface (Top surface of Single point cutting tool) and the reference plane or perpendicular surface of the workpiece. It plays a crucial role while designing cutting tools. It is the surface which allows the chip flow over them. It decides the shear angle at which chips will shear from

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

workpiece which further will responsible for the amount of heat generation while cutting, Cutting force, Cutting Power requirements, Tool life etc.

Rake angle can be classified into three types:

1. Positive Rake Angle
2. Zero Rake Angle
3. Negative Rake Angle

1. Positive Rake Angle:

If the Inclination of the Tool's top surface OR slope of the top surface is away from the cutting edge in the oppositedirection and at inner side, it is known as Positive Rake angle. It is usually made up of High speed steel material.

2. Zero Rake Angle:

When the Top surface of Cutting tool (Rake surface) is perpendicular to the Length of the cutting tool OR we can say, When the Rake surface of the tool coincides with the Reference plane of tool and makes No Angle with the reference plane, is described as Zero rake angle.

3. Negative Rake Angle:

If the Inclination of the Tool's top surface OR slope of the top surface is away from the cutting edge in the Same direction and at outer side, it is known as Negative Rake angle. It is usually made up of Carbide material to machinethe Harder material & to sustain the Higher Cutting Force.

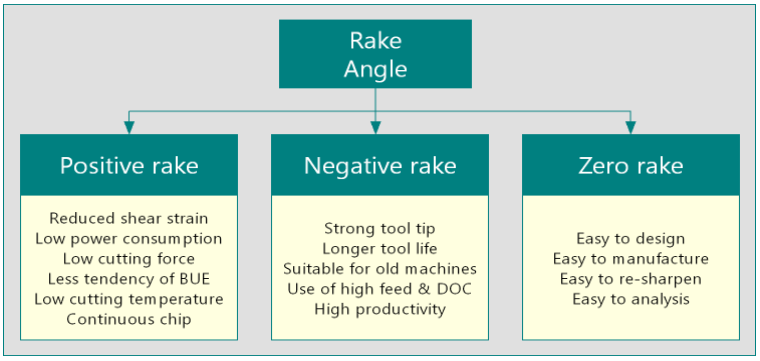


Figure 2: Flowchart of Different Types of Rake Angles

II. ANALYSIS

High-speed steel for the tool and steel 4340 for the workpiece are used in ANSYS simulations for any deformation operations. After selecting the relevant attributes, HSS is added to the library of Ansys workbench. The effect of the parameters like depth of cut, feed rate and cutting speed on cutting forces were also identified[2].

Simscale was used to create the mesh, which has 3.5K cells and 933 nodes.

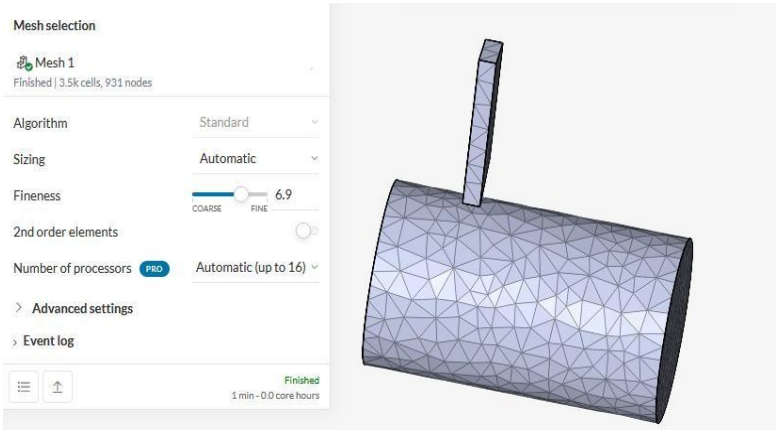


Figure 3: Mesh

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

Technological characteristics of machining process Several technological parameters are used to characterize cutting process. Various characteristics should be calculated like cutting power, strength values, heat and temperature and vibration[1]. Cutting process is characterized using Geometrical variables like: kinematic variables, cutting edge, feed, type ofchips, etc.

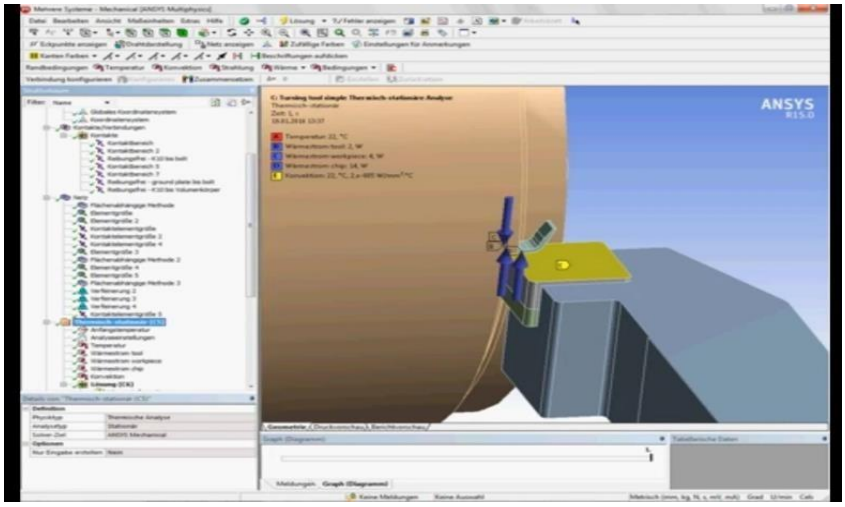


Figure 4: Cutting forces while Turning

The parameter calculation for turning of cylindrical surface

The parameter calculation for turning of cylindrical surface

Cutting speed, $v_c = \frac{\pi.D.N}{1000}$ m/min --------(1)

Feed rate - $v_f = \frac{f.N}{1000}$ m/min --------(2)

Speed of cutting process - $v_e = \sqrt{v_c^2 + v_f^2}$ m/min ---------- (3)

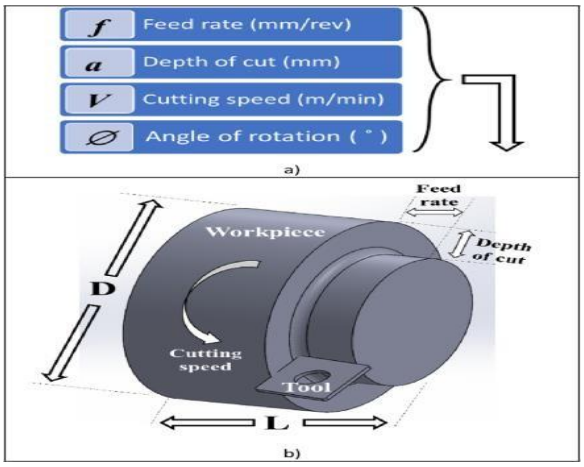


Figure 5: Consideration of Parameters of Tools & Workpiece

1. Positive Rake Angle (HSS Tool) - Aluminium Workpiece

We have used optimum data sets for the positive rake angle cutting tool:-

Back-rake angle = 12°; Side-rake angle =8°; End-Clearance angle = 8°; Side-Clearance angle = 7°; End-cutting edge angle = 16°; Side-cutting edge angle = 17°; Nose radius = 0.5mm.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

Table 1: Cutting Forces in Positive Rake Angle cutting tool

Positive Rake Angle (HSS Tool) - Aluminium Workpiece					
Positive Rake Angle	Rake Angle	Feed (mm.rev^{-1})	Fx - (N)	Fy - (N)	Fz - (N)
	12°	0.115	166.76	45.05	137.65
		0.150	245.155	59.708	151.036
		0.222	279.482	76.423	187.238

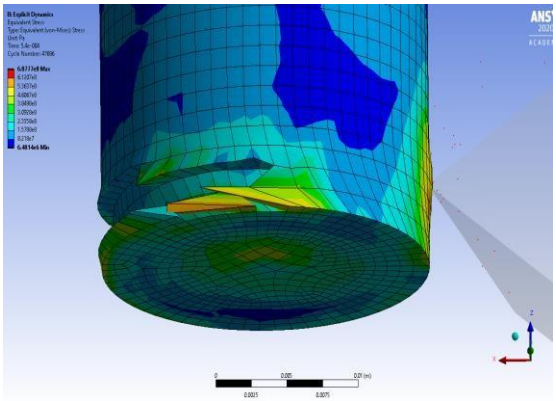


Figure 6: Cutting & Formation of chips in Turning

The cutting force is reduced by a positive rake angle. As a result, the cutting power required for the operation is reduced. If we use a positive rake angle then there are chances of continuous chip formation because chips are easily climbable over the rake surface of tool due to less shear angle between cutting chips and workpiece[5].

The cutting tool gets sharper as a result of the positive rake angle, which diminishes the cutting tool's strength and shortens its Tool's lifetime.

2. Zero Rake Angle (HSS Tool) - Brass Workpiece

We have used optimum data sets for the negative rake angle cutting tool:-x`

Back-rake angle = 0°; Side-rake angle =7°; End-Clearance angle = 7°; Side-Clearance angle = 7°;End-cutting edge angle = 15°; Side-cutting edge angle = 15°; Nose radius = 0.5mm.

Table 2: Cutting Forces in Zero Rake Angle cutting tool

Zero Rake Angle (Cemented Carbide Tool) - Titanium Workpiece					
Zero RakeAngle	Rake Angle	Feed (mm.rev^{-1})	Fx - (N)	Fy - (N)	Fz - (N)
	0°	0.115	290.56	229.59	211.15
		0.150	346.46	268.56	239.56
		0.222	418.26	318.79	285.63

- Neutral cutting power consumption happens. It is Easy to manufacture due to its simple Design. Cost of this type of cutting tool is cheaper.
- In this type of tool, we can see that there is a more chances of Formation of built-up edge.
- Efficiency of Cutting is less.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

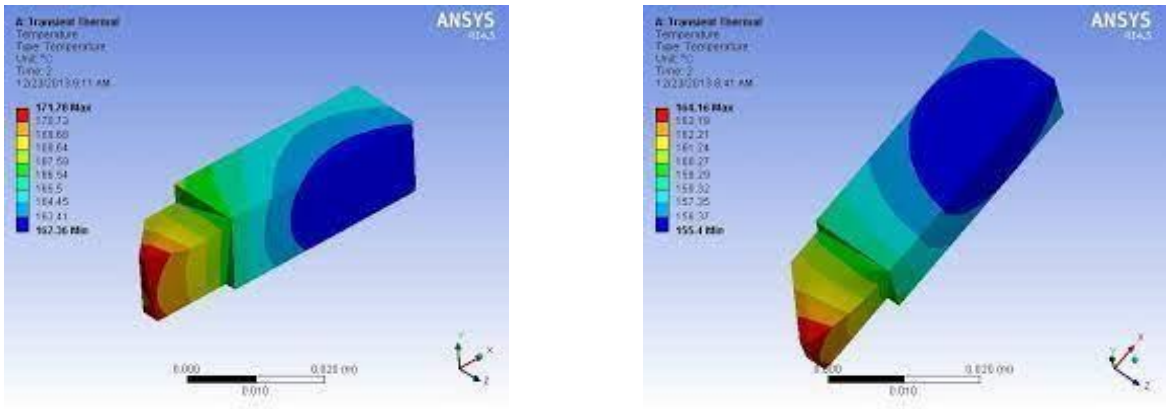


Figure 7: Temperature variation in Zero rake angle cutting tool

3. Negative Rake Angle (Cemented carbide tool)- Titanium Workpiece

We have used optimum data sets for the negative rake angle cutting tool:-

Back-rake angle = 18°; Side-rake angle =6°; End-Clearance angle = 5°; Side-Clearance angle = 6°; End-cutting edge angle = 14°; Side-cutting edge angle = 14°; Nose radius = 0.5mm.

Table 3: Cutting Forces in Negative Rake Angle cutting tool

Negative Rake Angle (Cemented Carbide Tool) – Titanium Workpiece					
Negative Rake Angle	Rake Angle	Feed (mm.rev^{-1})	Fx - (N)	Fy - (N)	Fz - (N)
	18°	0.115	79.85	45.51	60.26
		0.150	147.89	63.56	83.26
		0.222	238.96	108.47	130.47

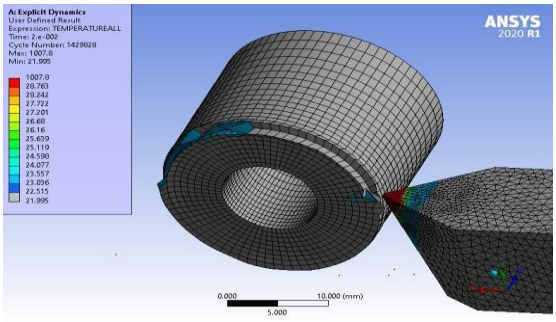


Figure 8: Temperature variation on positive rake angle cutting tool

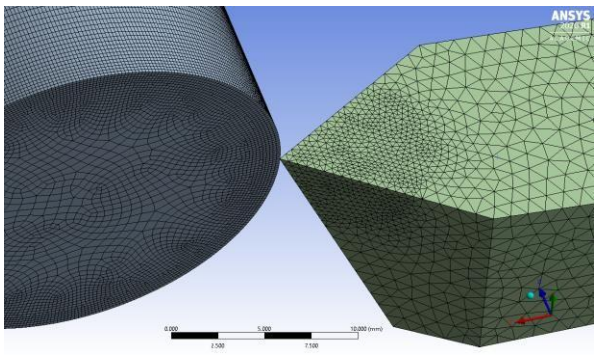


Figure 9: Mesh view

- Strength of Tool having Positive rake angle is more due to its Thicker cutting edge
- Tool life is More.
- Power required for Cutting is High because the negative rake angle makes the tool blunt.

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:05/May-2022 Impact Factor- 6.752 www.irjmets.com

- By using this type of tool the working surface temperature is increased due to friction, so coolingwater must be needed for the operation.
- This Type of tool is used for Machining Harder materials.

III. CONCLUSION

The results obtained in the study are summarized below.[6]

Table 4: Cutting Force (Fx) in X-direction for All Three Rake Angle cutting tool

Variation of Rake Angle & Feed Rate on Cutting Force (Fx (N))				
S. No.	Rake Angle	Feed Rate (mm/rev.)		
		0.115	0.150	0.222
1	12° (Positive Rake)	166.76	245.155	279.482
2	0° (Zero Rake)	290.56	346.46	418.26
3	18° (Negative Rake)	79.85	147.89	238.96

- F_c and F_f was increased, when Depth of cut and feed rate was increased. On the other hand, As Cutting speed was increased, Cutting forces was decreased whereas, it was noted that there was very less effect of cutting speed, as can be seen from the results[2].
- Fr was increased, when the feed rate was increased. However, a little change is observed in the value of Fr with the increasing depth of cut. This shows little variations in force distributions with the angle of approach of tool.
- It has been noted that Feed Force (Force in X- direction) was Greater than Tangential force(Force in Y-direction). And we also noted that radial force (Force in Z- direction) very small when it compares with other two force components.
- With the increase in feed rate, the cutting forces increases as expected.
- When we increase the rake angle from 0° from 18°, Cutting force will decrease.
- We have observed that Greatest Positive Rake angles lies from 12°to 16°.

IV. REFERENCES

[1] Zhang, Yancheng, J. C. Outeiro, and Tarek Mabrouki. "On the selection of Johnson-Cook constitutive model parameters for Ti-6Al-4 V using three types of numerical models of orthogonal cutting." Procedia Cirp 31 (2015): 112-117.

[2] Zhou, Lirong, Jianfeng Li, Fangyi Li, Gamini Mendis, and John W. Sutherland. "Optimization parameters for energy efficiency in end milling." Procedia CIRP 69 (2018): 312-317.

[3] Korkmaz, Mehmet Erdi, Nafiz Yaşar, and Mustafa Günay. "Numerical and experimental investigation of cutting forces in turning of Nimonic 80A superalloy." Engineering Science and Technology, an International Journal 23, no. 3 (2020): 664-673.

[4] Lapshin, Viktor, Denis Moiseev, and Valentin Minakov. "Diagnosing cutting tool wear after change of cutting forces during turning." In AIP Conference Proceedings, vol. 2188, no. 1, p. 030001. AIP Publishing LLC, 2019.

[5] Xie, Nan, Junfeng Zhou, and Beirong Zheng. "Selection of optimum turning parameters based on cooperative optimization of minimum energy consumption and high surface quality." Procedia CIRP 72 (2018): 1469-1474.Experiment measurement of cutting forces in the turning technology

[6] Glaa, Nawel, Kamel Mehdi, and Redouane Zitoune. "Numerical modeling and experimental analysis of thrust cutting force and torque in drilling process of titanium alloy Ti6Al4V." The International Journal of Advanced Manufacturing Technology 96, no. 5 (2018): 2815-2824.

[7] Kaymakci, M., Z. M. Kilic, and Y. Altintas. "Unified cutting force model for turning, boring, drilling and milling operations." International Journal of Machine Tools and Manufacture 54 (2012): 34-45.

[8] Hanief, M., M. F. Wani, and M. S. Charoo. "Modeling and prediction of cutting forces during theturning of

red brass (C23000) using ANN and regression analysis." Engineering science and technology, an international journal 20, no. 3 (2017): 1220-1226.

[9] Rao, C. J., D. Nageswara Rao, and P. Srihari. "Influence of cutting parameters on cutting forceand surface finish in turning operation." Procedia Engineering 64 (2013): 1405-1415.

[10] Thangarasu, S. K., S. Shankar, A. Tony Thomas, and G. Sridhar. "Prediction of Cutting Force inTurning Process-an Experimental Approach." In IOP conference series: materials science and engineering, vol. 310, no. 1, p. 012119. IOP Publishing, 2018.

[11] Cica, Djordje, Branislav Sredanovic, Gordana Lakic-Globocki, and Davorin Kramar. "Modeling of the cutting forces in turning process using various methods of cooling and lubricating: an artificial intelligence approach." Advances in Mechanical Engineering 5 (2013): 798597.

[12] Chen, Xi, Dinghua Zhang, and Qi Wang. "Cutting Force Transition Model Considering the Influence of Tool System by Using Standard Test Table." Sensors 21, no. 4 (2021): 1340.

[13] Patole, Pralhad B., and Vivek V. Kulkarni. "Prediction of surface roughness and cutting forceunder MQL turning of AISI 4340 with nano fluid by using response surface methodology." Manufacturing Review 5 (2018)

[14] Shankar, Subramaniam, S. K. Thangarasu, T. Mohanraj, and D. S. Pravien. "Prediction of cuttingforce in turning process: An experimental and fuzzy approach." Journal of Intelligent & Fuzzy Systems 28, no. 4 (2015): 1785-1793.

[15] Yousefi, Sasan, and Mehdi Zohoor. "Effect of cutting parameters on the dimensional accuracy and surface finish in the hard turning of MDN250 steel with cubic boron nitride tool, for developing a knowledged base expert system." International Journal of Mechanical and Materials Engineering 14, no. 1 (2019).

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360597771

PROCEEDINGS OF ITR INTERNATIONAL CONFERENCE Kolkata, India Organized by Date of Event

Conference Paper · April 2022

CITATIONS
0

READS
7

1 author:



Debottam Bhowmik
Delhi Technological University
4 PUBLICATIONS **0** CITATIONS

SEE PROFILE

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

Fitting Regression Model and Optimization of Cutting Parameters In Order To Achieve Economical Machining Of a Cryogenic Soaked Magnesium Alloy View project

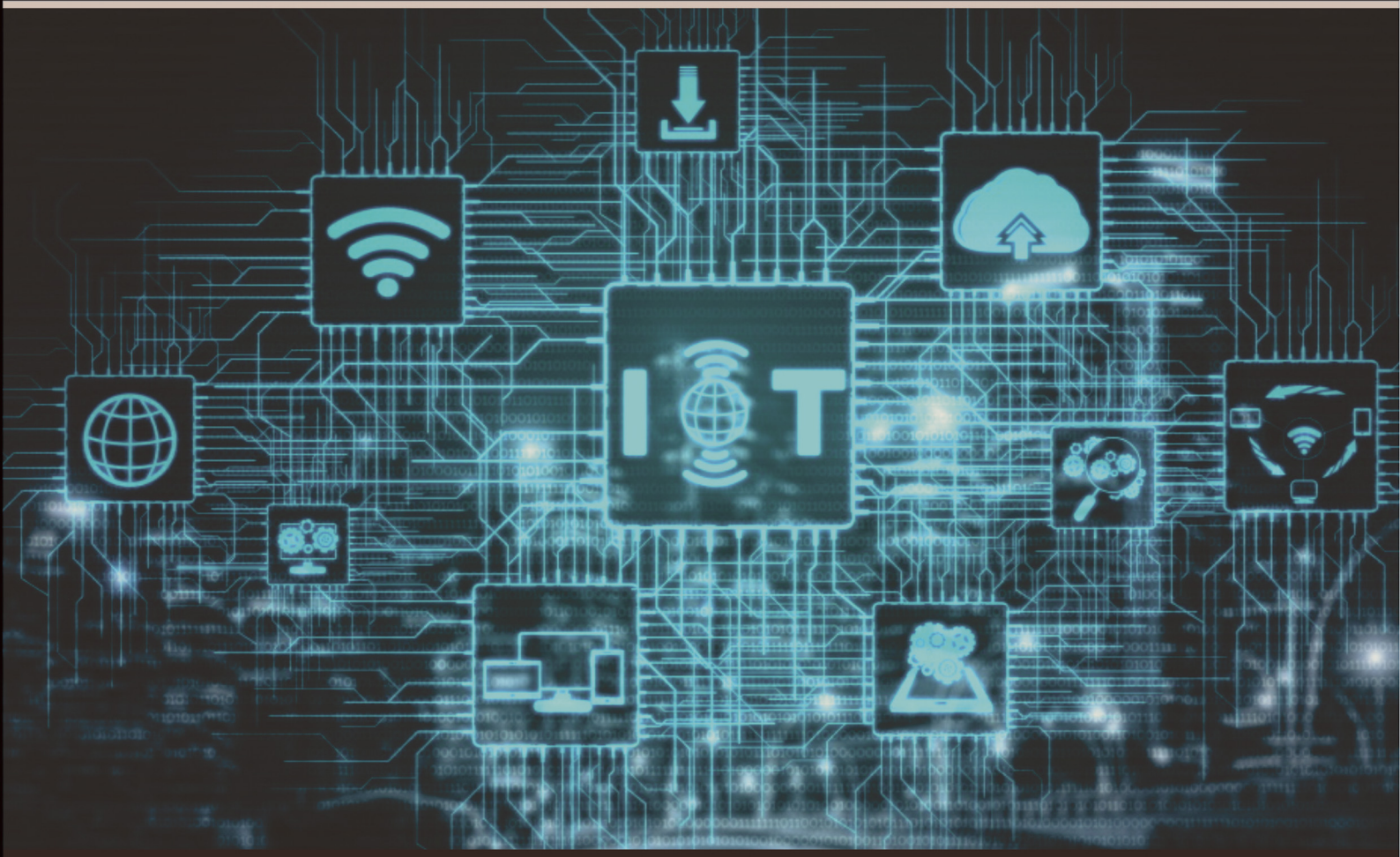
All content following this page was uploaded by Debottam Bhowmik on 14 May 2022.

The user has requested enhancement of the downloaded file.

PROCEEDINGS OF

ITR

INTERNATIONAL CONFERENCE



17TH APRIL, 2022 | VENUE: KOLKATA, INDIA

Association With:

PROCEEDINGS OF

ITR

INTERNATIONAL CONFERENCE

KOLKATA, INDIA

Organized by

Date of Event

17th April, 2022

In Association with

Corporate Address
INSTITUTE OF RESEARCH AND JOURNALS
Plot No- 30, Dharma Vihar, Khandagiri, Bhubaneswar, Odisha, India
Mail: info@iraj.in, **www.iraj.in**

Publisher: **Institute for Technology and Research (ITRESEARCH)**

© 2022, ITR International Conference, Kolkata, India

ISBN: 978-93-90150-31-1

Edn: 389

No part of this book can be reproduced in any form or by any means without prior written permission of the publisher.

Disclaimer: Authors have ensured sincerely that all the information given in this book is accurate, true, comprehensive, and correct right from the time it has been brought in writing. However, the publishers, the editors, and the authors are not to be held responsible for any kind of omission or error that might appear later on, or for any injury, damage, loss, or financial concerns that might arise as consequences of using the book.

Type set & Printed by:

Institute for Technology and Research (ITRESEARCH)
Bhubaneswar, India

About Institute of Research and Journals:

Institute of Research and Journals (IRAJ) is an advanced Non-profit technological forum under Peoples Empowerment Trust, for the Researchers & Scholars "to promote the progress of Science and Technology" by displaying their knowledge in the vicinity of Science and Technology for the service of mankind and the advancement of general welfare.

Objective of IRAJ:

- To provide a world class platform to researchers to share the research findings by organizing International/National Conferences.
- To use the research output of the conference in the class room for the benefits of the students.
- To encourage researchers to identify significant research issues in identified areas, in the field of Science, Engineering, Technology and Management.
- To help dissemination of their work through publications in a journal or in the form of conference proceedings or books.
- To help them in getting feedback on their research work for improving the same and making them more relevant and meaningful, through collective efforts.
- To encourage regional and international communication and collaboration; promote professional interaction and lifelong learning; recognize outstanding contributions of individuals and organizations; encourage scholar researchers to pursue studies and careers in circuit branches and its applications.
- To set up, establish, maintain and manage centers of excellence for the study of /on related subjects and discipline and also to run self supporting projects for the benefit of needy persons, irrespective of their caste, creed or religion.

About ITR:

Institute For Technology and Research (ITResearch) is one of the world's largest Technological and Scientific research and evaluation organization. The aim of this advanced technological forum for the researchers & scholars is to exhibit their knowledge in the area of science and technology. ITResearch is founded in 2012 as a not-for-profit organization and registered under a trust. We conduct our work with strict independence. We are standing tall in the field of research & technology, our objective is to provide a platform which will encourage and support scholars, researchers & professionals to carry and accomplish their research. ITResearch is an open association of excellent Scientists, Academicians, Professionals, Engineers and Scholars.

Conference Committee

Program Chair:

Dr. P. Suresh
M.E, Ph.D. Professor and Controller of Examinations,
Karpagam College of Engineering,
Coimbatore, India.

Conference Manager:
Mr. Bijan Kumar Barik

Conference Convener:

Miss. Suchismita Sahoo
Mob: +91-8895188971

Miss. Sumita Nayak
Mob: +91-8280047516

Publication Head:
Mr. Manas Ranjan Prusty

INTERNATIONAL ADVISORY MEMBERS

Prof. Goodarz Ahmadi,
Professor, Mechanical and Aeronautical Engineering, Clarkson University, USA

Dr Chi Hieu Le,
Senior Lecturer, University of Greenwich. Kent ME4 4TB. United Kingdom

PROF. (ER.) Anand Nayyar
Department of Computer Applications & I.T.KCL Institute of Management and Technology, Jalandhar
G.T. Road, Jalandhar-144001,Punjab, India.

Prof. R. M. Khaire,
Professor, Dept. Of Elex. and Telecommunication, B, V University, India

Dr.P.Suresh,
Professor, Karpagam College of Engineering, Coimbatore,Tamilnadu

Mark Leeson
Associate Professor (Reader)
Area of Expertise: nanoscale communications, evolutionary algorithms, network coding and communication systems

Dr. P. K. Agarwal
Professor, Deptt. of Civil Engineering, MANIT Bhopal ,Ph. D: IIT Kanpur
M.E: Civil Engg.IIT Roorkee, Membership: Indian Road Congress (IRC), Institute of Urban Transport (IUT)

Shahriar Shahbazpanahi
Islamic Azad University,
Department of Civil Engineering, Sanandaj, Kurdistan, Iran, PhD (Structural Engineering),
University Putra Malaysia, Malaysia, 2009-Present

Harun Bin Sarip
Head of Research and InnovationDept, UniKL-MICET
Doctorate: Université de La Rochelle, France
Member : International Society of Pharmaceutical Engineer, Singapore Chapter

Dr. Md. Al-Amin Bhuiyan
Associate Professor
Dept. of Computer Engineering, King Faisal University
Al Ahssa 31982, Saudi Arabia

Prof. (Er.) Anand nayyar
Department of Computer Applications & I.T.
KCL Institute of Management and Technology, Jalandhar
G.T. Road, Jalandhar-144001, Punjab, India

Prof. Aleksandr Cariow
institution or Company: West Pomeranian University of Technology, Szczecin

Dr. P. K. Agarwal
Professor, Deptt. of Civil Engineering, MANIT Bhopal ,Ph. D: IIT Kanpur
M.E: Civil Engg.IIT Roorkee, Membership: Indian Road Congress (IRC), Institute of Urban Transport (IUT)

Dr. VPS Naidu
Principal Scientist & Assoc. Prof., MSDF Lab, FMCD
CSIR - National Aerospace Laboratories, Bangalore, India

Mr. P. Sita Rama Reddy
Chief Scientist ,Mineral Processing Department, CSIR - Institute of Minerals & Materials Technology
Bhubaneswar, India, M.Tech. (Chem. Engg., IIT, KGP)

Dr.P.C.Srikanth,
Professor & Head, E&C Dept, Malnad College of Engineering, Karnataka
Senior Member IEEE, Secretary IEEE Photonics Society,
M.Tech: IIT, Kanpur, Ph.D: In IISc Photonics lab

Prof. Lalit Kumar Awasthi,
Professor, Department of Computer Science & Engineering
National Institute of Technology(NIT-Hamirpur),
PhD, IIT, Roorkee, M. Tech, IIT, Delhi

Dr. Chandra Mohan V.P.
Assistant Professor, Dept. of Mech. Engg., NIT Warangal,
Warangal. Ph.D : Indian Institute of Technology(IIT),Delhi
M.B.A: Alagappa University

Prof. I.Suneetha,
Associate Professor, Dept. of ECE, AITS, Tirupati, India

Dr.S. Chandra Mohan Reddy,
Assistant Professor (SG) & Head,Dept. of Electronics &
Communication Engineering, JNTUA College of Engineering,
Pulivendula, Ph.D,J.N.T. University Anantapur, Anantapuramu

Gurudatt Anil Kulkarni,
I/C HOD E&TC Department, MARATHWADA MITRA
MANDAL'S POLYTECHNIC

Pasuluri Bindu Swetha
Dept. Of ECE, Stanley college of Engineering & Technology for
Women, Hyderabad, India

★★★

TABLE OF CONTENTS

Sl No	TITLES AND AUTHORS	Page No.
01.	**Fitting Regression Model and Plotting Contour Plots to Investigate Critical Cutting Parameters for Achieving Economical Machining Output of a Cryogenic Soaked Magnesium Alloy** ➢ *Debottam Bhowmik, N.Yuvaraj*	**1-10**
02.	**Home Office Communication on the Instagram Social Network during Covid-19** ➢ *Lucie Pilařová, Tereza Balcarová, Ladislav Pilař*	**11-14**
03.	**A Keyword Spotting Approach using Agglomerative Hierarchical Clustering Method** ➢ *Sushma S N, Sharada B*	**15-20**
04.	**Automated Tool for Scheduling Cases in Judiciary** ➢ *Manish Kumar, Prashanth M V*	**21-26**
05.	**A Comparative Study of Machine Learning Algorithm to Predict Used Car Prices in India with EDA** ➢ *Rakesh Sharma, K.T Thomas*	**27-33**
06.	**The Effects of Job Retention Schemes on Employment Preservation during the COVID-19 Epidemic in Euro Area Countries** ➢ *Anton Rop*	**34-42**
07.	**Recycling Spent Coffee Waste for Innovative Thermo-Plasters: Development of a Prototype** ➢ *Manfredi Saeli, Tiziana Campisi*	**43-50**
08.	**The Antecedents of Customer Loyalty in the Café Industry: Physical environment, Service quality, Food quality, Beverage quality** ➢ *Ryo Sakiyama*	**51-55**
09.	**Digital Preservation: A Need of Tomorrow** ➢ *Sumeet, Nisha*	**56-58**
10.	**Does Personality Matter? A Study of Retail Banks in Hong Kong** ➢ *Ho, Shirie Pui Shan, Chow Matthew Yau Choi*	**59-61**

★★★

EDITORIAL

It is my proud privilege to welcome you all to the ITR International Conference at Kolkata, India. I am happy to see the papers from all part of the world and some of the best paper published in this proceedings. This proceeding brings out the various Research papers from diverse areas of Science, Engineering, Technology and Management. This platform is intended to provide a platform for researchers, educators and professionals to present their discoveries and innovative practice and to explore future trends and applications in the field Science and Engineering. However, this conference will also provide a forum for dissemination of knowledge on both theoretical and applied research on the above said area with an ultimate aim to bridge the gap between these coherent disciplines of knowledge. Thus the forum accelerates the trend of development of technology for next generation. Our goal is to make the Conference proceedings useful and interesting to audiences involved in research in these areas, as well as to those involved in design, implementation and operation, to achieve the goal.

I once again give thanks to the Institute of Research and Journals, ITR & TheIIER for organizing this event in Kolkata, India. I am sure the contributions by the authors shall add value to the research community. I also thank all the International Advisory members and Reviewers for making this event a Successful one.

Editor-In-Chief

Dr. P. Suresh
M.E, Ph.D. Professor and Controller of Examinations,
Karpagam College of Engineering,
Coimbatore, India.

★★★

FITTING REGRESSION MODEL AND PLOTTING CONTOUR PLOTS TO INVESTIGATE CRITICAL CUTTING PARAMETERS FOR ACHIEVING ECONOMICAL MACHINING OUTPUT OF A CRYOGENIC SOAKED MAGNESIUM ALLOY

[1]DEBOTTAM BHOWMIK, [2]N.YUVARAJ

[1]M.Tech Student, Department of Mechanical Engineering, Delhi Technological University, Delhi, India
[2]Assistant Professor, Department of Mechanical Engineering, Delhi Technological University, Delhi, India
E-Mail: [1]kingdeb720@gmail.com, [2]yuvraj@dce.ac.in

Abstract - The objective of this work is to identify the critical cutting parameter out of speed, feed, depth of cut and approach angle of the cutting tool by fitting linear regression model between the output variable and the input variable along with analysis of variance method to compute contribution of each of these parameters on the quality of machined output. The response output variables taken up for this analysis were surface roughness, material removal rate, tool wear rate and machining time as all these factors governs the quality of the machined product and hence relates to economics of machining. Contour plots have been drawn to study about interactions between the input cutting parameters and its impact on the output variables. From the results it was observed that surface roughness is mostly affected by feed closely followed by approach angle of cutting tool whereas material removal rate, tool wear rate and machining time were mostly affected by depth of cut. It was also observed that cryogenic soaking suitably increases the hardness of the material thus making it handier for machining purpose.

Keywords - Economical Machining, Regression modelling, CNC Turning, Cryogenic Soaking, Magnesium alloy

I. INTRODUCTION & MOTIVATION

The ultimate objective for any production or manufacturing unit is to produce a product which is economical in terms of overall cost of production. This motive of attaining economical machining is achieved by analysing the machining process and identifying critical cutting parameters which has definite influence on the machining operation so that it can be optimized which will eventually leads to obtain an economical machining output. Thus identification of critical of parameters plays a paramount role. In present times, many industries are giving more importance in reducing the overall cost of production and in this regard reduction in weight of the work material has become necessary which in turn reduces the power consumption required for machining and hence leads to reduction in overall costs. Thus usage of light weight materials has garnered deep interest in the industries. This project deals with the turning of Mg AZ31B alloy for the above stated reasons. Among the many available light weight materials, aluminium alloys are vastly used in various industrial applications. From its utility in coke canes to aircraft fuselage, aluminium alloys due to its cheaper cost, ease of fabrication and lighter weight enjoys most popularity among the other materials. But there are some drawbacks with regards to aluminium alloys like it results in abrasive and scratchy machined surface due to its poor wear resistance which becomes detrimental for using it in case of precise engineering output as required in making propeller blades of aircraft, in making of transmission casing used in helicopters where there is need for even lighter material. Along with this, the poor biocompatibility of aluminium when it is used in case of food packaging and cooking utensils might link to Alzheimer's disease as stated by Christopher exley et al (2014)[1]. To solve the above mentioned problems of aluminium alloys demand for new options emerged and magnesium alloys proved to be a viable alternative. A literature review has been undertaken in order to obtain useful information regarding machining of magnesium alloy. It involves identifications and articulations of linkage between the literature and our topic of inquiry. Although the format of the reviewing literature can be altered by using different types of criteria, the purpose of the research remains unchanged. Viswanathan et al. (2014)[2] aimed to optimize cutting parameters such as cutting speed, feed rate, and depth of cut while dry turning an AZ91D Mg alloy. Among the multiple factors listed above, feed rate was shown to be the most impacting parameter. Adel et al (2018)[3] built an artificial neutral network using matlab to predict surface roughness over minimal machining time at prime machining cost. The ANN model was able to predict with 1.35 percent accuracy. Dutta & Narala et al (2021)[4] successfully optimized turning parameters for newly developed AM Mg alloy using Taguchi technique. It was concluded that depth of cut bears maximum influence on cutting forces and feed has maximum influence of surface roughness. Danish et al (2019)[5] investigated the influence of cryogenic treatment while performing turning operation of AZ31C alloy in order to analyse surface integrity parameters. It concluded that cryogenic machining improves surface quality, corrosion resistance and reduces ignition risk. Bruschi et al (2018)[6] have experimentally performed the turning operation on

AZ31 alloy under three different cutting conditions i.e. dry cutting, wet cutting and cryogenic cooling. It was concluded by them that the process with cryogenic treatment results into machined surface with improved corrosion resistance. Dutta et al (2020) [7]explained that even nose radius has an important role in deciding machined surface quality by performing turning operation on a magnesium work specimen. The result indicated that larger the nose radius better will be the surface produced. Song et al (2013)[8] stated that AZ91D Mg alloys has the potential to be the implant biomaterial as it is harmless for human being, due to the fact that magnesium alloys are biodegradable in fluid of human body and Mg^{+2} helps in quick growth of bone tissue . Chakraborty Banerjee et al (2019)[9] investigated various types of material to analyse about their natural biocompatibility which includes magnesium, titanium, cobalt chromium alloys as well as stainless steel. Among these materials, it was observed that magnesium alloys were found to be having best potential for biomaterial implants as it has lower corrosion rate in human blood fluid as compared to others.

According to the above-mentioned literature study, the majority of researchers have undertaken optimization analysis using feed, speed, and depth of cut as likely input factors. Only a few of them have used cutting tool geometry as an input parameter, and the majority have not explored it. Similarly, the output variable has primarily focused on surface roughness, with only a handful taking into account other variables.

Based on a comprehensive analysis of the literature, it was determined to include the aspect of cutting tool geometry in this project, and therefore, in addition to speed, feed, and depth of cut, the approach angle of the cutting tool would be one of the input parameters. We will use surface roughness, material removal rate, tool wear rate, and machining time as output parameters to examine economical machining, which is directly related to the quality of the product produced after machining.

II. METHODOLOGY

The experiment has been performed on a single piece magnesium AZ31B alloy. The specimen was of cylindrical bar shape with diameter 20 mm and the length of the specimen was taken as 300 mm. Table 1 shows the chemical composition of the alloy

COMPONENTS	% BY WEIGHT
Mg	97
Al	2.50 – 3.50
Zn	0.60 – 1.40
Man	0.20
Si	0.10
Cu	0.05
Ca	0.04
Fe	0.005
Ni	0.005

Table 1 Chemical composition of AZ31B alloy

The magnesium AZ31B has been turned by using carbide inserts having angles 35˚, 55˚, 80˚. Table 2 shows the specification the cutting tool.

CUTTING TOOL PARAMETERS	SPECIFICATIONS
ANSI Number	CNMG120408
Brand Name	W-Walter
Class	Turning
Code	CNMG120408-MV5 WPV20
Grade	MV5 WPV20
Fixing hole diameter	5.16 mm
Insert style	CNMG
Insert size	432
Material	Carbide
Corner radius	0.8 mm
Insert thickness	4.76 mm
Type	Turning insert

Table 2 Key specifications of watler cutting tool

The experimental work has been carried out on CNC LT-16 XL having machine number 381 using CNMG inserts having tool corner radius 0.8 mm. The CNC system used is Fanue Oi Mate-TD and the CNC package used is Beeta 8i s (feed motor- 2 Nos), Beeta 6i (spindle), with SVPM-11 i.



Figure 1: CNC LT-16 XL

The surface roughness measurement of the nine turned out specimens with regard to this project was done by using a portable surface roughness tester named as Mitutoyo surftest SJ210 series.



Figure 2 : Mitutoyo Surftest SJ210 Series

Regression model has been used to identify the the critical parameter among the speed, feed, depth of cut and approach angle of the cutting tool. Regression modelling is basically a statistical approach that defines mathematical relationship between one or more independent variables with as many numbers of dependent variables. In this project, we will be dealing with linear regression modelling where the dependent and independent variables are related to one another through a linear equation as stated below.

$$y = a + bx + e$$

Where, y = dependent variable
x = independent variable
a = intercept value of y when x is 0
b = regression coefficient
e = error of the estimate

The most important table among all the tables provided after performing regression analysis is the model summary table. Some of the key elements which will feature in the model summary table are pointed out as below:

- S: It is the deviation between fitted value and data value. Lower the value of S better will be the model however lower value of S not always provide correct scenario
- R-Sq: It is used to explain the percentage variation of response by the model. Higher of R-Sq provides better fit to the model. Its value lies between 0 to 100%
- R-Sq (adj): It is used to compare models that have different number of predictors. It helps in determining which of the parameters might not be helpful in improving the model and hence that can be discarded.

III. REGRESSION MODELLING

ANALYSIS OF SURFACE ROUGHNESS

The table given below represents the obtained experimental values of surface roughness along with the different levels of input parameters used in this project.

Run	Cutting Speed (rpm)	Feed (mm/rev)	Depth of Cut (mm)	Approach Angle (°)	Surface Roughness (µm)
1	800	0.2	0.2	35	5.084
2	800	0.3	0.3	55	7.290
3	800	0.4	0.4	80	3.448
4	1000	0.2	0.3	80	1.801
5	1000	0.3	0.4	35	7.729
6	1000	0.4	0.2	55	12.335
7	1200	0.2	0.4	55	3.353
8	1200	0.3	0.2	80	2.833
9	1200	0.4	0.3	35	15.037

Table 3 Experimental values of surface roughness

In order to fit a linear regression equation, it must satisfy the assumption that all the concerned data are having normal distribution. Hence normality test must be performed and it is indicated by plotting a probability plot for the given data and the P-value for probability plot must be more than 0.05 in order pass

the normality test. The figure given below shows the probability plot for surface roughness.

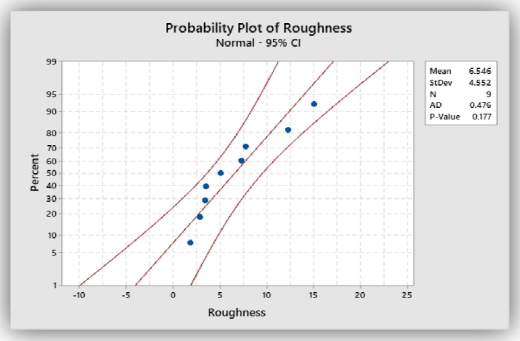


Figure 3 Probability plot of surface roughness

Thus the regression equation was obtained as

$$Ra = 3.03 + 0.00450*v + 34.30*f - 9.54*d - 0.1486*\theta$$

Where, Ra = surface roughness (µm)

v = cutting speed (rpm)

f = feed (mm/rev)

d = depth of cut (mm)

θ = approach angle (˚)

The first important table which we obtained was the Coefficients table listed below as table 4

Term	Coefficient	SE coefficient	T- Value	P- Value	VIF
Constant	3.03	6.05	0.50	0.642	
Cutting Speed	0.00450	0.00428	1.05	0.352	1.00
Feed	34.30	8.55	4.01	0.016	1.00
Depth of Cut	-9.54	8.55	-1.12	0.327	1.00
Approach Angle	-0.1486	0.0379	-3.92	0.017	1.00

Table 4 Coefficients table of surface roughness

Thus we can conclude that that feed has the maximum influence on the surface roughness having a P-value of 0.016 which is closely followed by approach angle of the cutting tool as the second most influencing parameter with a P-value of 0.017. All the input parameters have a VIF value of 1.00 which shows that the regression coefficients are not correlated among themselves.

The next important table which we obtain as a part of the regression analysis is model summary table. The table 5 listed below represents the model summary.

S	R-sq	R-sq (adj)
2.09471	89.41 %	78.83 %

Table 5 Model summary

From the above table, it can be concluded that the regression model for surface roughness has a R^2 value of 89.41 %.

The next important table which we obtain was that of ANOVA analysis.

Source	DF	Adj SS	Adj MS	F-value	P-value	%contribution
Cutting Speed	1	4.862	4.862	1.11	0.352	3.28
Feed	1	70.603	70.603	16.09	0.016	47.63
Depth of Cut	1	5.457	5.457	1.24	0.327	3.68
Approach Angle	1	67.321	67.321	15.34	0.017	45.41
Error	4	17.551	4.388			
Total	8	165.794				100

Table 6 ANOVA table for regression model of surface roughness

The above results can also be described by using a pareto chart for standardized effect with surface roughness being the response. The figure is given below.

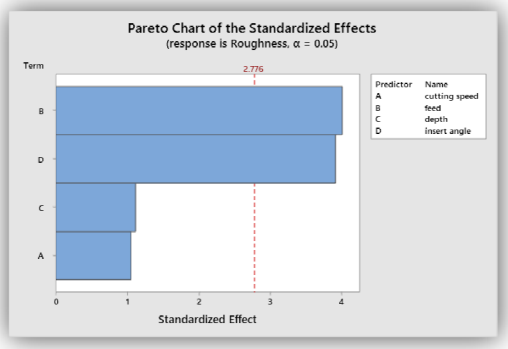


Figure 4 Pareto chart of standardized effects for surface roughness

The last outcome which we obtain after performing a regression analysis was residual plots for surface roughness. This is a combination of four plots namely histogram, normal probability plot, versus fits and versus order. The residual plots for surface roughness are shown in the figure below.

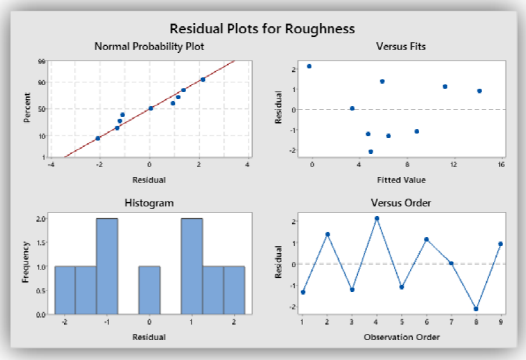


Figure 5 Residual plots for surface roughness

We can observe that the histogram has no outliers and is somewhat skewed. The normal probability plot is approximately linear which means it is in consistent with a normal distribution. The residual versus fitted value shows random pattern which suggest that residual has constant variance. The residual versus observation order shows random pattern which means that the data was collected and can be used to obtain non-random error.

ANALYSIS OF MATERIAL REMOVAL RATE

The table given below represents the obtained experimental values of material removal rate along with the different levels of input parameters used in this project.

Run	Cutting Speed (rpm)	Feed (mm/rev)	Depth of Cut (mm)	Approach Angle (°)	Material Removal Rate (mm^3/min)
1	800	0.2	0.2	35	1234.91
2	800	0.3	0.3	55	2077.10
3	800	0.4	0.4	80	3323.36
4	1000	0.2	0.3	80	2435.22
5	1000	0.3	0.4	35	3037.48
6	1000	0.4	0.2	55	1793.56
7	1200	0.2	0.4	55	3138.73
8	1200	0.3	0.2	80	2206.92
9	1200	0.4	0.3	35	3869.67

Table 7 Experimental values of material removal rate

In order to fit a linear regression equation, it must satisfy the assumption that all the concerned data are having normal distribution. Hence normality test must be performed and it is indicated by plotting a probability plot.
The figure given below shows the probability plot for material removal rate.

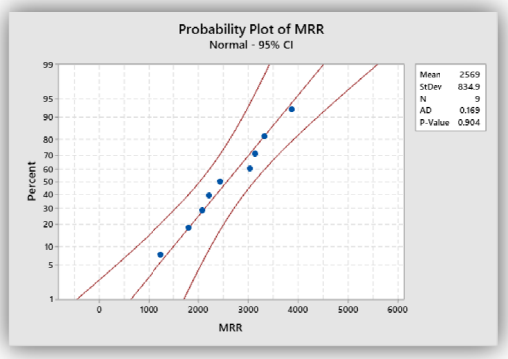


Figure 6 Probability plot of material removal rate

Thus the regression equation was obtained as

$$MRR = -2761 + 2.150*v + 3630*f + 7107*d - 0.73*\theta$$

Where, MRR = material removal rate (mm^3/min)
v = cutting speed (rpm)
f = feed (mm/rev)
d = depth of cut (mm)
θ = approach angle (˚)

The first important table which we obtained was the Coefficients table, has been listed below as table 8

Term	Coefficient	SE coefficient	T-Value	P-Value	VIF
Constant	-2761	1159	-2.38	0.076	
Cutting Speed	2.150	0.820	2.62	0.059	1.00
Feed	3630	1639	2.21	0.091	1.00
Depth of Cut	7107	1639	4.34	0.012	1.00
Approach Angle	-0.73	7.27	-0.10	0.925	1.00

Table 8 Coefficients table of material removal rate

We can conclude that that depth of cut has the maximum influence on the material removal rate having a P-value of 0.012 followed by cutting speed as the second most influencing parameter with a P-value of 0.059. All the input parameters have a VIF value of 1.00 which shows that the regression coefficients are not correlated among themselves. The next important table which we obtain as a part of the regression analysis is model summary table. The table 9 listed below represents the model summary.

S	R-sq	R-sq (adj)
401.555	88.43 %	76.87

Table 9 Model summary

From the above table, it can be concluded that the regression model for material removal rate has a R^2 value of 88.43 %.

The next important table which we obtain after performing the regression analysis was that of analysis of variance also called as ANOVA table as listed below.

Source	DF	Adj SS	Adj MS	F-value	P-value	%contribution
Cutting Speed	1	1109353	1109353	6.88	0.059	22.49
Feed	1	790417	790417	4.90	0.091	16.03
Depth of Cut	1	3030546	3030546	18.79	0.012	61.45
Approach Angle	1	1633	1633	0.01	0.925	0.033
Error	4	644985	161246			
Total	8	5576934				100

Table 10 ANOVA table for regression model of material removal rate

The above results can also be described by using a pareto chart for standardized effect with material removal rate being the response.

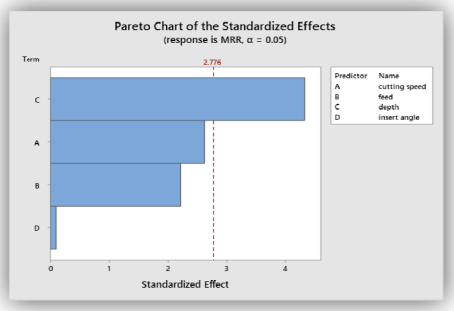


Figure 7 Pareto chart of standardized effects for material removal rate

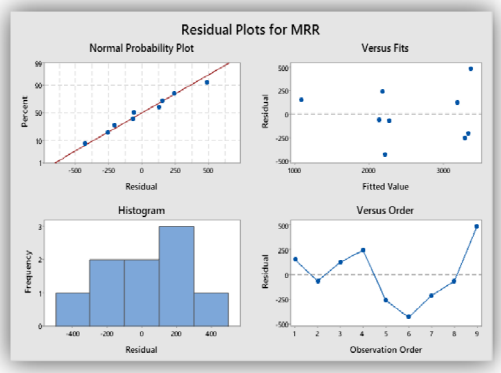


Figure 8 Residual plots for material removal rate

The last outcome which we obtain after performing a regression analysis was residual plots for material removal rate. The residual plots for material removal rate are shown in the figure below.

We can observe that the histogram has no outliers and is skewed. The normal probability plot is approximately linear which means it is in consistent with a normal distribution. The residual versus fitted value shows random pattern which suggest that residual has constant variance. The residual versus observation order shows random pattern which means that the data was collected and can be used to obtain non-random error.

ANALYSIS OF TOOL WEAR RATE

The table given below represents the obtained experimental values of tool wear rate along with the different levels of input parameters used in this project.

Run	Cutting Speed (rpm)	Feed (mm/rev)	Depth of Cut (mm)	Approach Angle (°)	Tool Wear Rate (mm^3/min)
1	800	0.2	0.2	35	1.748
2	800	0.3	0.3	55	2.352
3	800	0.4	0.4	80	6.021
4	1000	0.2	0.3	80	3.309
5	1000	0.3	0.4	35	4.127
6	1000	0.4	0.2	55	3.044
7	1200	0.2	0.4	55	4.976
8	1200	0.3	0.2	80	3.001
9	1200	0.4	0.3	35	4.382

Table 11 Experimental values of tool wear rate

In order to fit a linear regression equation, it must satisfy the assumption that all the concerned data are having normal distribution. Hence normality test must be performed and it is indicated by plotting a probability plot.

The figure given below shows the probability plot for tool wear rate.

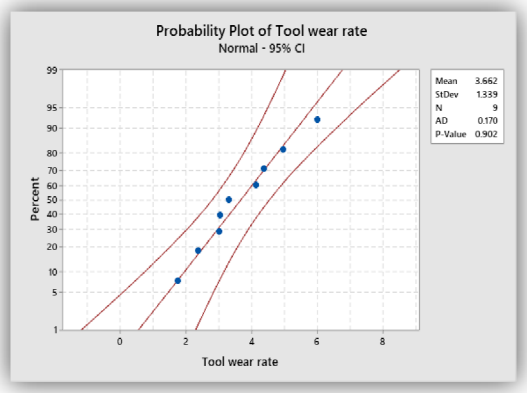


Figure 9 Probability plot of tool wear rate

Thus the regression equation was obtained as

TWR = -4.47 + 0.00186*v + 5.69*f + 12.22*d + 0.0158*θ

Where, TWR = Tool wear rate (mm^3/min)
v = cutting speed (rpm)
f = feed (mm/rev)
d = depth of cut (mm)
θ = approach angle (˚)

The first important table which we obtained was the Coefficients table; it has been listed below as table 12

Term	Coefficient	SE coefficient	T-Value	P-Value	VIF
Constant	-4.47	1.96	-2.28	0.085	
Cutting Speed	0.00186	0.00139	1.34	0.251	1.00
Feed	5.69	2.78	2.05	0.110	1.00
Depth of Cut	12.22	2.78	4.40	0.012	1.00
Approach Angle	0.0158	0.0123	1.28	0.269	1.00

Table 12 Coefficients table of tool wear rate

We can conclude that that depth of cut has the maximum influence on the tool wear rate having a P-value of 0.012 followed by feed as the second most influencing parameter with a P-value of 0.110. All the input parameters have a VIF value of 1.00 which shows that the regression coefficients are not correlated among themselves.

The next important table which we obtain as a part of the regression analysis is model summary table. The table 13 listed below represents the model summary.

S	R-sq	R-sq (adj)
0.680566	87.09%	74.18%

Table 13 Model summary

From the above table, it can be concluded that the regression model for tool wear rate has a R^2 value of 87.09 %.

The next important table which we obtain was that of analysis of variance also called as ANOVA table.

Source	DF	Adj SS	Adj MS	F-value	P-value	%contribution
Cutting Speed	1	0.8348	0.8348	1.80	0.251	6.68
Feed	1	1.9426	1.9426	4.19	0.110	15.55
Depth of Cut	1	8.9573	8.9573	19.34	0.012	71.67
Approach Angle	1	0.7618	0.7618	1.64	0.269	6.1
Error	4	1.8527	0.4632			
Total	8	14.3491				100

Table 14 ANOVA table for regression model of tool wear rate

The above results can also be described by using a pareto chart for standardized effect with tool wear rate being the response. The figure provided below showing the requisite pareto chart.

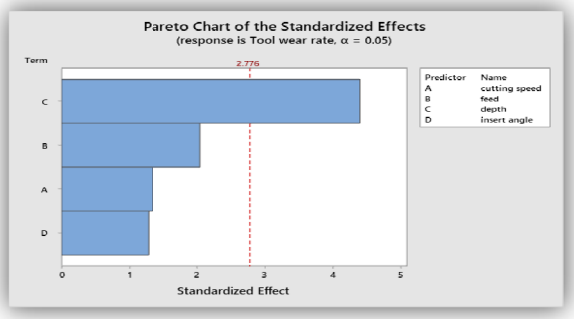


Figure 10 Pareto chart of standardized effects for tool wear rate

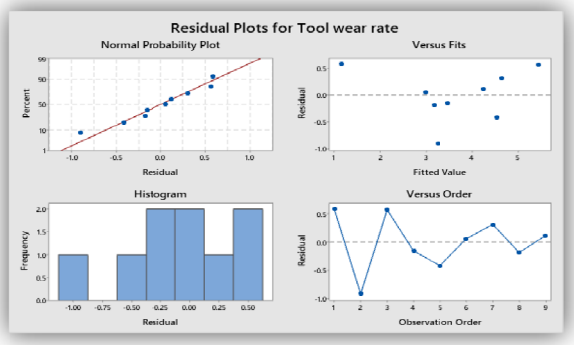


Figure 11 Residual plots for tool wear rate

At the end we obtain residual plots for tool wear rate. The residual plots for tool wear rate are shown in the figure below.

We can observe that the histogram has no outliers and is somewhat skewed. The normal probability plot is approximately linear which means it is in consistent with a normal distribution. The residual versus fitted value shows random pattern which suggest that residual has constant variance. The residual versus observation order shows random pattern which means that the data was collected and can be used to obtain non-random error.

ANALYSIS OF MACHINING TIME

The table given below represents the obtained experimental values of machining time along with the different levels of input parameters used in this project.

Run	Cutting Speed (rpm)	Feed (mm/rev)	Depth of Cut (mm)	Approach Angle (˚)	Machining Time (min)
1	800	0.2	0.2	35	1.83
2	800	0.3	0.3	55	1.36
3	800	0.4	0.4	80	0.85
4	1000	0.2	0.3	80	1.16
5	1000	0.3	0.4	35	0.93
6	1000	0.4	0.2	55	1.26
7	1200	0.2	0.4	55	0.90
8	1200	0.3	0.2	80	1.28
9	1200	0.4	0.3	35	0.73

Table 15 Experimental values of machining time

In order to fit a linear regression equation, it must satisfy the assumption that all the concerned data are having normal distribution. Hence normality test must be performed and it is indicated by plotting a probability plot.

The figure given below shows the probability plot for machining time.

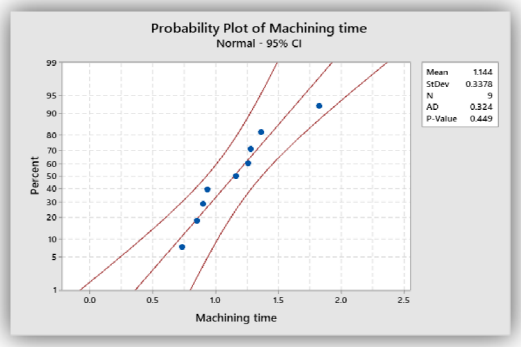


Figure 12 Probability plot of machining time

The regression equation was obtained as

MTM = 3.3544 – 0.000942*v -1.750*f – 2.817*d – 0.00155*θ

Where, MTM = Machining time (min)
v = cutting speed (rpm)
f = feed (mm/rev)
d = depth of cut (mm)
θ = approach angle (˚)

The first important table which we obtained was the Coefficients table; it has been listed below as table 16

Term	Coefficient	SE coefficient	T-Value	P-Value	VIF
Constant	3.544	0.261	13.57	0.000	
Cutting Speed	-0.000942	0.000185	-5.10	0.007	1.00
Feed	-1.750	0.369	-4.74	0.009	1.00
Depth of Cut	-2.817	0.369	-7.63	0.002	1.00
Approach Angle	-0.00155	0.00164	-0.94	0.399	1.00

Table 16 Coefficients table of machining time

We can conclude that that depth of cut has the maximum influence on the machining time having a P-value of 0.002 followed by cutting speed as the second most influencing parameter with a P-value of 0.007. The feed has much significance with P-value 0.009. All the input parameters have a VIF value of 1.00 which shows that the regression coefficients are not correlated among themselves.

The table 17 listed below represents the model summary for machining time.

S	R-sq	R-sq (adj)
0.0904776	96.41 %	92.82 %

Table 17 Model summary

From the above table, it can be concluded that the regression model for machining time has a R^2 value of 96.41 %.

Source	DF	Adj SS	Adj MS	F-value	P-value	%contribution
Cutting Speed	1	0.212817	0.212817	26.00	0.007	24.18
Feed	1	0.183750	0.183750	22.45	0.009	20.88
Depth of Cut	1	0.476017	0.476017	58.15	0.002	54.11
Approach Angle	1	0.007294	0.007294	0.89	0.399	0.83
Error	4	0.032745	0.007294			
Total	8	0.912622				100

Table 18 ANOVA table for regression model of machining time

The next important table which we obtain after performing the regression analysis was that of analysis of variance also called as ANOVA table.

The above results can also be described by using a pareto chart for standardized effect with tool wear rate being the response. The figure provided below showing the requisite pareto chart.

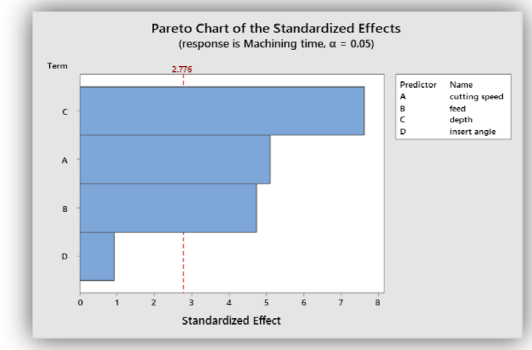


Figure 13 Pareto chart of standardized effects for machining time

At last we obtain residual plots which were shown in the figure given below.

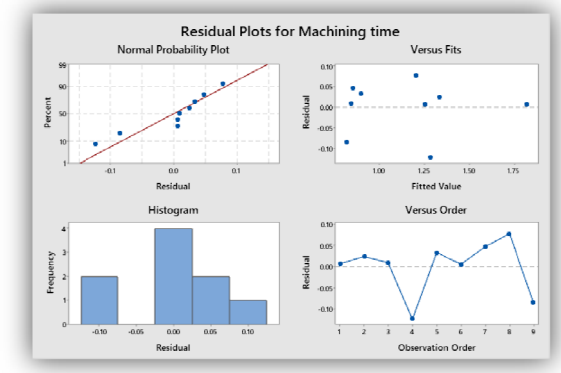


Figure 14 Residual plots for machining time

We can observe that the histogram has no outliers and is somewhat skewed. The normal probability plot is approximately linear which means it is in consistent with a normal distribution. The residual versus fitted value shows random pattern which suggest that residual has constant variance. The residual versus observation order shows random pattern which means that the data was collected and can be used to obtain non-random error.

IV. HARDNESS ANALYSIS

As we know that cryogenic soaking helps in improving mechanical properties of the component, an experimental investigation has been performed to check the hardness of the component after the soaking duration and to compare it with the hardness of the component before cryogenic soaking. It was done by Vickers hardness testing machine and a load of 100g has been applied on the specimen in 3 different trials for about 10 seconds of time.

Trial	Hardness before soaking (HV)	Hardness after soaking (HV)	% improvement in hardness
1st Trial	82	116	
2nd Trial	86	109	33.73
3rd Trial	81	108	
Average	83	111	

Table 19 Hardness of AZ31B alloy before and after cryogenic soaking

Thus an improvement of about 33.73 % has been observed after soaking the Mg AZ31B alloy for about 30 minutes in this project work.

The trend obtained in the improvement of hardness before and after cryogenic soaking of the component is similar to the results obtained by Kaveh Meshinchi Asl et al (2009)[10].

V. CONCLUSION

- Regression analysis has been performed in order to evaluate mathematical relation between the output parameters like surface roughness, material removal rate, tool wear rate, machining time and the input parameters like cutting speed, feed, depth of cut and approach angle of cutting tool.
- The model summary concluded that regression model for surface roughness has a R^2 value of 89.41 %, material removal rate with R^2 value of about 88.43 %. The R^2 value of regression model for tool wear rate and machining time was 87.09 % and 96.41 % respectively.
- As per ANOVA analysis for regression model, surface roughness was heavily influenced by feed with 47.63% closely followed by approach angle of cutting tool with 45.41 % and cutting speed along with depth of cut has a contribution of about 3.28 % and 3.68 % respectively.
- As per ANOVA analysis for regression model, depth of cut has maximum influence on material removal rate with 61.45 % and is followed by cutting speed with 22.49%. The feed has 16.03% contribution and approach angle of cutting tool has a miniscule contribution of 0.033%.
- As per ANOVA analysis for regression model, depth of cut has maximum influence on tool wear rate with 71.67 % and is followed by feed with 15.55%. The cutting speed has 6.68% contribution and approach angle of cutting tool has a contribution of about 6.1%.
- As per ANOVA analysis for regression model, depth of cut has maximum influence on machining time with 54.11 % and is followed cutting speed with 24.18%. The feed has 20.88%

contribution and approach angle of cutting tool has a miniscule contribution of about 0.83%.

- In order to investigate about interactions between the input parameters and its impact on the output parameter contour plots has been drawn for each of the output parameters like surface roughness, material removal rate, tool wear rate and machining time against the two input parameters combination out of the available input parameters.
- It was found that hardness after cryogenic soaking has been increased by 33.73 %.

REFERENCES

[1] C. Exley, "What is the risk of aluminium as a neurotoxin?," Expert Rev. Neurother., vol. 14, no. 6, pp. 589–591, 2014, doi: 10.1586/14737175.2014.915745.

[2] M. Kumar, "Optimization of Machining Parameters of Aluminum Alloy Using Taguchi Approach," SSRN Electron. J., no. December, pp. 10–14, 2020, doi: 10.2139/ssrn.3554237.

[3] A. T. Abbas, D. Y. Pimenov, I. N. Erdakov, M. A. Taha, M. S. Soliman, and M. M. El Rayes, "ANN surface roughness optimization of AZ61 magnesium alloy finish turning: Minimum machining times at prime machining costs," Materials (Basel)., vol. 11, no. 5, 2018, doi: 10.3390/ma11050808.

[4] S. Dutta and S. Kumar Reddy Narala, "Optimizing turning parameters in the machining of AM alloy using Taguchi methodology," Meas. J. Int. Meas. Confed., vol. 169, p. 108340, 2021, doi: 10.1016/j.measurement.2020.108340.

[5] M. Danish, T. L. Ginta, K. Habib, A. M. Abdul Rani, and B. B. Saha, "Effect of Cryogenic Cooling on the Heat Transfer during Turning of AZ31C Magnesium Alloy," Heat Transf. Eng., vol. 40, no. 12, pp. 1023–1032, 2019, doi: 10.1080/01457632.2018.1450345.

[6] S. Bruschi, R. Bertolini, A. Ghiotti, E. Savio, W. Guo, and R. Shivpuri, "Machining-induced surface transformations of magnesium alloys to enhance corrosion resistance in human-like environment," CIRP Ann., vol. 67, no. 1, pp. 579–582, 2018, doi: 10.1016/j.cirp.2018.04.040.

[7] S. Dutta and S. K. R. Narala, "Effect of tool nose radius in turning of novel Mg alloy," Mater. Today Proc., vol. 38, no. xxxx, pp. 2675–2679, 2020, doi: 10.1016/j.matpr.2020.08.334.

[8] Y. W. Song, D. Y. Shan, and E. H. Han, "Electrodeposition of hydroxyapatite coating on AZ91D magnesium alloy for biomaterial application," Mater. Lett., vol. 62, no. 17–18, pp. 3276–3279, 2008, doi: 10.1016/j.matlet.2008.02.048.

[9] P. C. Banerjee, S. Al-Saadi, L. Choudhary, S. E. Harandi, and R. Singh, "Magnesium implants: Prospects and challenges," Materials (Basel)., vol. 12, no. 1, pp. 1–21, 2019, doi: 10.3390/ma12010136.

[10] K. M. Asl, A. Tari, and F. Khomamizadeh, "Effect of deep cryogenic treatment on microstructure, creep and wear behaviors of AZ91 magnesium alloy," Mater. Sci. Eng. A, vol. 523, no. 1–2, pp. 27–31, 2009, doi: 10.1016/j.msea.2009.06.003.

★★★

IRAJ INTERNATIONAL JOURNALS



IJMPE

www.ijmpe.iraj.in

International Journal of Mechanical and Product Engineering
ISSN(P):2320-2092
ISSN(e):2321-2071



IJEEDC

www.ijeedc.iraj.in

International Journal of Electrical, Electronics and Data Communication
ISSN(P):2320-2084
ISSN(e):2321–2950



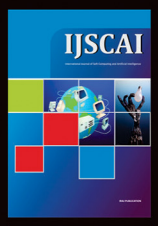
IJACEN

www.ijacen.iraj.in

International Journal of Advance Computational Engineering and Networking
ISSN(e):2320-2106
ISSN(P):2320-2106

Indexing Partners

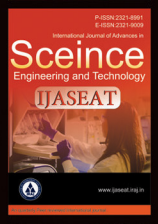




International Journal of Soft Computing And Artificial Intelligence (IJSCAI)
ISSN(P):2321–404X
ISSN(e):2321–4384
www.ijscai.iraj.in



International Journal of Advances in Computer Science and Cloud Computing (IJACSCC)
ISSN(P):2321–4058
ISSN(e):2321–4392
www.ijacscc.iraj.in



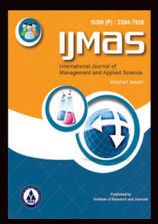
International Journal of Advances in Science, Engineering and Technology(IJASEAT)
ISSN(P):2321-8991
ISSN(e):2321–9009
www.ijaseat.iraj.in



International Journal of Industrial Electronics and Electrical Engineering(IJIEEE).
ISSN(P):2347-6982
ISSN(e):2349-204X
www.ijieee.iraj.in



International Journal of Advances in Mechanical and Civil Engineering(IJAMCE)
ISSN(P):2394-2827
www.ijamce.iraj.in



International Journal of Management and Applied Science (IJMAS)
ISSN(P):2394-7926
www.ijmas.iraj.in



International Journal of Advance in Electronics and Computer Science (IJAECS)
ISSN(P):2393-2835
www.ijaecs.iraj.in

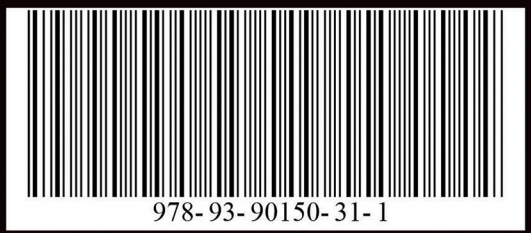
IRAJ Journals Listed in University Library
MIT, University of California Berkeley, Stanford, Cambridge, Oxford, Harvard
Visit for Upcoming Conferences - **www.itresearch.org.in**

ICA International Conference Alert
www.conferencealert.com
ALLCONFERENCEALERT
www.allconferencealert.com
Conference Gallery .com

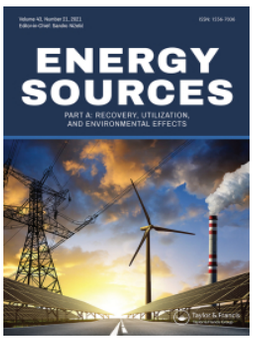
Digital Xplore

This book will be available online at DigitalXplore
www.digitalxplore.org

ISBN



View publication stats



Energy Sources, Part A: Recovery, Utilization, and Environmental Effects

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ueso20

Process optimization of conventional steam distillation system for peppermint oil extraction

Ravi Kant & Anil Kumar

To cite this article: Ravi Kant & Anil Kumar (2022) Process optimization of conventional steam distillation system for peppermint oil extraction, Energy Sources, Part A: Recovery, Utilization, and Environmental Effects, 44:2, 3960-3980, DOI: 10.1080/15567036.2022.2069886

To link to this article: https://doi.org/10.1080/15567036.2022.2069886

Published online: 17 May 2022.

Submit your article to this journal

Article views: 18

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=ueso20

ENERGY SOURCES, PART A: RECOVERY, UTILIZATION, AND ENVIRONMENTAL EFFECTS
2022, VOL. 44, NO. 2, 3960–3980
https://doi.org/10.1080/15567036.2022.2069886

Taylor & Francis
Taylor & Francis Group

Check for updates

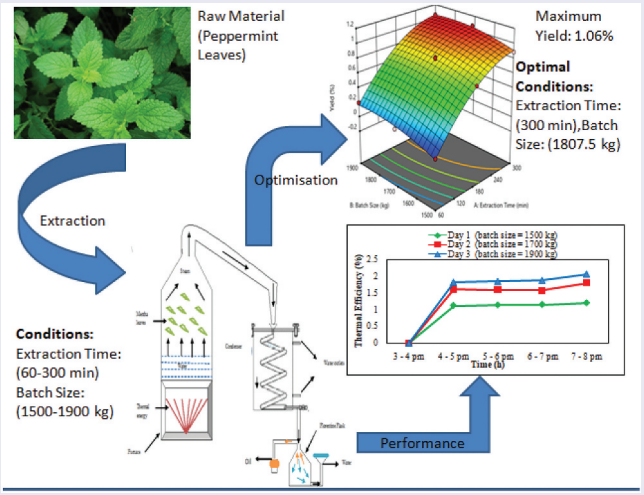
Process optimization of conventional steam distillation system for peppermint oil extraction

Ravi Kant [a] and Anil Kumar [a,b]

[a]Department of Mechanical Engineering, Delhi Technological University, Delhi, India; [b]Centre for Energy and Environment, Delhi Technological University, Delhi, India

ABSTRACT

This study aims to optimize peppermint oil extraction process using response surface methodology (RSM). Experiments were performed for different batch sizes (1500, 1700, and 1900 kg). System performance was evaluated in terms of thermal efficiency, productivity, essential oil yield (EOY), extraction efficiency, and CO_2 emissions. Total energy consumption for 1500, 1700, and 1900 kg batch sizes are 1379.46, 1336.58, and 1286.56 kWh, while CO_2 emissions are 1104, 1069, and 1028 kg, respectively. Maximum hourly system productivity, cumulative productivity, maximum hourly thermal efficiency, maximum extraction efficiency, and maximum EOY were obtained as 8.3 L/h, 20 L, 2%, 41.5%, and 0.461% (w/w), respectively, for 1900 kg. Average increase in total productivity, extraction efficiency, and EOY was 49.25%, 1%, and 26%, increasing in batch size from 1500 kg to 1900 kg. Extraction time and batch size have important effect ($p < 0.001$) on the extraction yield. Optimal process parameters are identified 300 min of extraction time and 1807.5 kg of batch size. The highest oil yield was 1.0614% under optimal conditions. Experiment was further performed for 1807.5 kg of batch size and the yield was found as 1.04621. Predicted and experimental results were found in good consistency with percentage deviation of 1.436. Therefore, this is helpful in conducting further research on optimizing process parameters for conventional distillation system of peppermint oil extraction.



ARTICLE HISTORY
Received 10 January 2022
Revised 4 April 2022
Accepted 13 April 2022

KEYWORDS
Peppermint; distillation; performance; essential oil yield; optimization; RSM

CONTACT Anil Kumar anilkumar76@dtu.ac.in Department of Mechanical Engineering, Delhi Technological University, 110 042 Delhi, India

© 2022 Taylor & Francis Group, LLC

Introduction

Herbal medicines are used by the majority of the population (80%) to address their health requirements. Therefore, medicinal and aromatic plants are essential since herbal medicines are made from their products. Approximately 21000 species are utilized as medicinal plants as per WHO records. Essential oils (EOs) are extracted from these medicinal plants by distillation process (Kant and Kumar 2021; Radwan et al. 2020b). These oils are used in food industry as flavoring agent, as fragrances in cosmetic products, and in the medical sector for functional use (Radwan et al. 2020b, 2020a). Peppermint (Mentha piperita L.) is a seasonal medicinal herb that belongs to a Lamiaceae (Labiatae) family. It is a Mediterranean native species cultivated world-wide for use in flavoring, fragrance, medicine, and pharmaceutical products (Iscan et al. 2002). It is widely used for nervous system and digestive disorder problems. The peppermint plant is cultivated in areas where the soil can hold water. All commercial mint varieties are sterile seeds propagated from underground rootstocks (stolons). The rootstocks are formed by the plants that already exist. Owing to the quick deterioration of rootstock due to heat and dryness, they cannot be stored for more than a few days (Loolaie et al. 2017). Menthol and menthone are the main constituents of peppermint oil (PO) and its chemical composition varies with geographic area, plant maturity, and processing conditions (Riachi and De maria, 2015). Figure 1 depicts the utilization of peppermint oil as a medical requirement.

Many processes are used for EO extraction from peppermint leaves, such as steam distillation (SD) (Radwan et al. 2020a), hydro-distillation (HD) (Memarzadeh et al. 2020), solvent extraction (Fadhil et al. 2020), supercritical fluid extraction (Peterson et al. 2006; Sodeifian et al. 2017), and subcritical extraction liquid (Kubatova et al. 2001). Mostly used processes are SD or HD (Belhachat et al. 2018; Liu et al. 2019). In SD, thermal energy of the steam or water is used to evaporate the oil from the leaves, and then the mixture of oil and water vaporizes condensed in the condenser.

The yield is affected by extraction parameters in any extraction method. The response surface methodology (RSM) is the optimization study that is widely used to optimize the extraction parameters of any process. The existing products and unit operations are designed, analyzed, and optimized using RSM. Its use reduces the time, number of experiments, initial

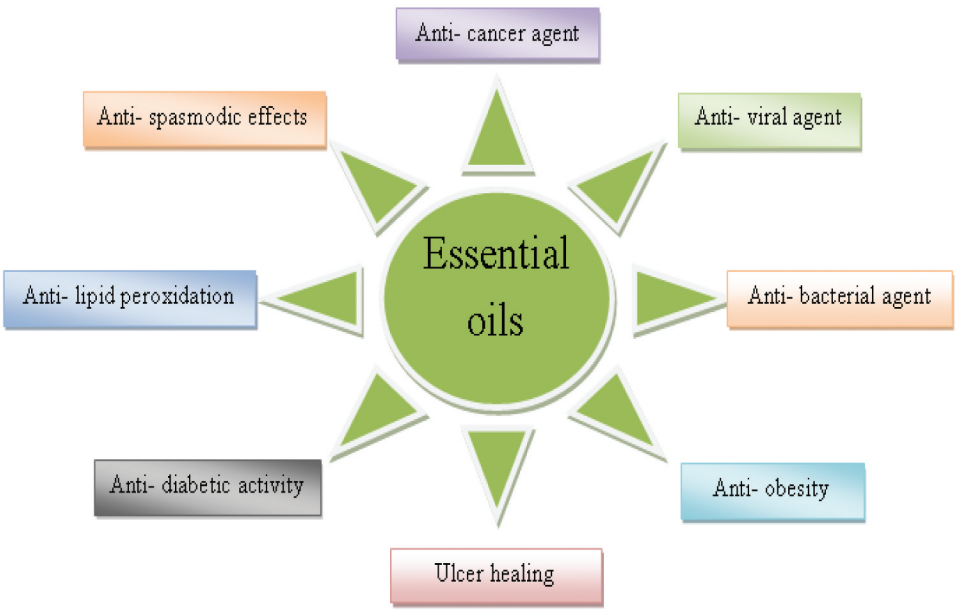


Figure 1. Peppermint oil uses as health benefits

investment, and energy compared to others (Belhachat et. 2018). Numerous studies have been carried out on performance and optimization of different EO extraction methods. Radwan et al. developed the SD system for peppermint (Mentha pipereta L.) and lavender oil. They studied the effects of boiler inlet water mass flow rate and batch size of Peppermint plants on productivity, extraction efficiency, EO yield, and power requirements. System productivity, essential oil yield (EOY), and extraction efficiency increase at higher inlet mass flow rate in boiler and larger batch size; however, energy requirement decreases (Radwan et al. 2020a, 2020b). Munir et al. designed and developed a solar distillery to extract EO from the aromatic and medicinal plant and calculated system efficiency (33.21%) and thermal energy requirement (1.548 kW) for distillation process. It was concluded that 3.5 kWh energy was required to process 10 kg batch (Munir et al. 2014). Afjal et al. developed an advanced hybrid solar distillation system (HSDS) to extract the EO from the eucalyptus and peppermint. EOY of eucalyptus and peppermint was estimated as 0.59% and 0.40%, respectively. In addition, biomass system was also coupled with the distillation system to supplement the system during extremely cold conditions or seasonal climatic conditions in a typical year (Afjal et al. 2017). Gavahian and Chu (2018) designed and developed an ohmic accelerated SD system for lavender oil and compared it with conventional steam distillation (CSD). Yield of the proposed system was 3.3–3.8% more than CSD system; nevertheless, the energy consumed for the same amount of lavender was 0.5 kWh lower than CSD. Kulturel and Tarhan experimented on a solar distillation system coupled with parabolic compound collectors and determined the system performance at different meteorological conditions. It resulted in 26–40 mL for 5 kg mint daily productivity in plant (Kulturel and Tarhan (2016). Yu Liu et al. used RSM to optimize the extraction parameters of ultrasonic-assisted HD to extract oil from Iberis amara seeds. Optimal extraction yield was obtained at an extraction time (ET) of 240 min, ultrasonic power (UP) of 75 W, and solvent-to-sample ratio (SSR) of 6.7 mL/g (Liu et al. 2019). Belhachat et al. (2018) used HD for essential extraction from ripe berries and employed RSM to optimize the extraction parameters. The optimal extraction conditions were a 60-min ET, a 60 W UP, and 1 g: 12 mL solid–liquid ratio . Rajeshwaran et al. (2020) used solvent extraction technique to extract oil from *Acacia nilotica* plant and optimize the input parameters by RSM. The maximum oil yield was obtained at a reaction temperature of 70°C, extraction time of 5.99 h, and a solvent-to-seed ratio of 5:1. Shekhar et al. (2021) used soxhlet-assisted solvent extraction technique to extract oil from rain tree seeds and used RSM to optimize the extraction process parameters. Process parameters were considered as ET, powder weight, and volume of solvent. Maximum yield of 11.15% was obtained at optimum condition of process parameter. Tran et al. (2018) optimized the microwave-assisted HD process's parameter using RSM. Basil plant was used for EO extraction in this study. The optimum conditions were obtained as 97 min ET, 3.2:1 liquid-to-material ratio, and 430 W microwave power. Sodeifian et al. (2019) extracted Diospyros lotus seed oil by conventional method and optimized the operating parameters using RSM. Optimal yield was obtained at 61°C temperature, 8.75 h extraction time, and 19.25 mL/g solvent to solid ratio.

Various studies on EO extraction from aromatic and medicinal plants are reviewed. Performance analysis and optimization of the conventional distillation system (biomass based) for oil extraction from peppermint in Indian region have not been done so far. In this work, performance analysis and optimization of extraction yield of peppermint oil using RSM of conventional distillation system were done, as it is useful for energy-efficient operation. Performance of the system was evaluated in terms of EOY, extraction efficiency, productivity, and thermal efficiency. RSM is a very important optimization tool since its use reduces the time, number of experiments, initial investment, and energy consumption compared to others. Extraction time and plant batch size are extraction parameters that were optimized in this study. Design expert 11.0 version with face-centered central composite

design (CCD) investigated the effects of extraction time and plant batch size on the extraction yield. The existing products and unit operations are designed, analyzed, and optimized using RSM.

Materials and methods

CSD system is installed at Birpura village, Jhansi, India (25° 36 N′18.40″E). Experiments are performed for one complete cycle during 3 days in November 2020 for various batch sizes of the entire plant (1500 kg, 1700 kg, and 1900 kg).

Sample preparation

Oil is extracted from peppermint (Mentha) plant using a conventional distillation process. High water holding capacity soil is required to cultivate peppermint plants. Peppermint plants are cultivated. Panida herbicide is used to reduce weed in the field. Mixture of Mentha power and urea fertilizer is used to grow fast peppermint plants and increase oil content. This crop is generally harvested in November in India. Peppermint leaves are collected after harvesting using farm machinery and transported from field to distillation unit by tractor and trolley. A digital weighing machine is used to weigh peppermint leaves. Peppermint plants are supplied with the help of conveyor in the distillation unit after weighing. Table 1 illustrates the parameters of the peppermint plant.

Conventional steam distillation system

Conventional distillation system mainly consists of four parts named as furnace, boiler, distillation unit, and condenser. Approximately 1500 kg, 1700 kg, and 1900 kg batch size of peppermint plants were weighed and placed in distillation unit for three cycles, respectively. Distillation unit consists of two net, one at the bottom, and other at the middle. Packed bed of leaves is prepared by putting peppermint leaves on these nets. Wood of mango tree was used as a fuel in furnace. Heat generated in furnace was transferred to boiler. Water in the boiler gets heated up and reached to boiling point temperature. Generated steam then passed through packed bed of peppermint in distillation unit. Peppermint leaves releases oil at a steam temperature of 250–350°C. Distillation unit was connected to condenser through a connection hose. Vapors of oil and water are then transferred to condenser through connection hose to complete the condensation process. Florentine flask was used to separate the oil and water. Schematic and experimental setup for conventional distillation system is shown in Figures 2 and 3, respectively. Description and working of all components used in the SD process are discussed in Table 2.

Table 1. Parameters of Peppermint plant (Radwan et al. 2020a, 2020b)

Sr. No.	Properties	Values
1.	Diameter of stem (mm)	2.37
2.	Length of stem (cm)	15–20
3.	Mass of plant (g)	3–5
4.	Number of leaves per plant	20–40
5.	Moisture content of leaves (%)	83.52
6.	Thickness of leaves (mm)	0.25
7.	Mass of whole Peppermint plant (g)	5

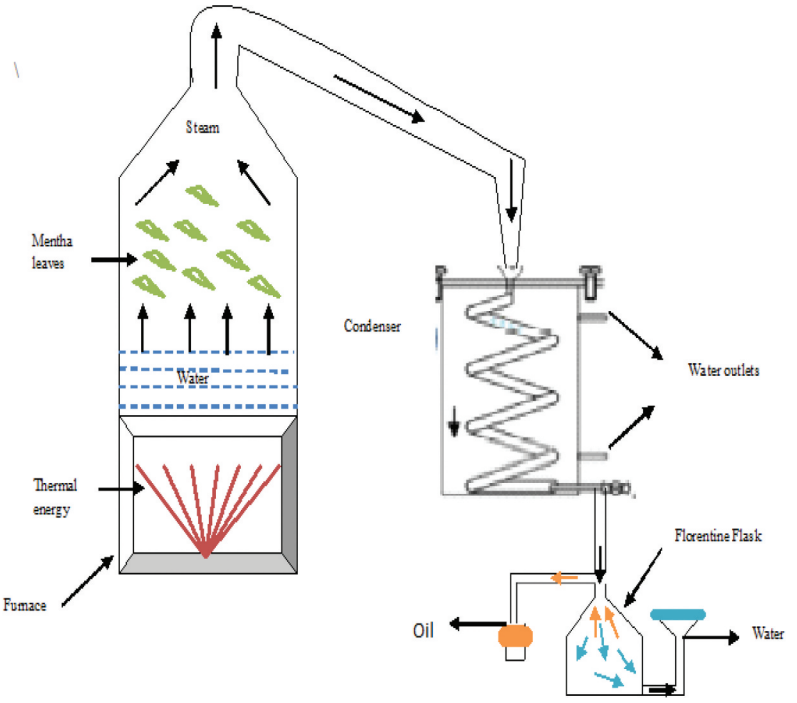


Figure 2. Schematic of conventional distillation system

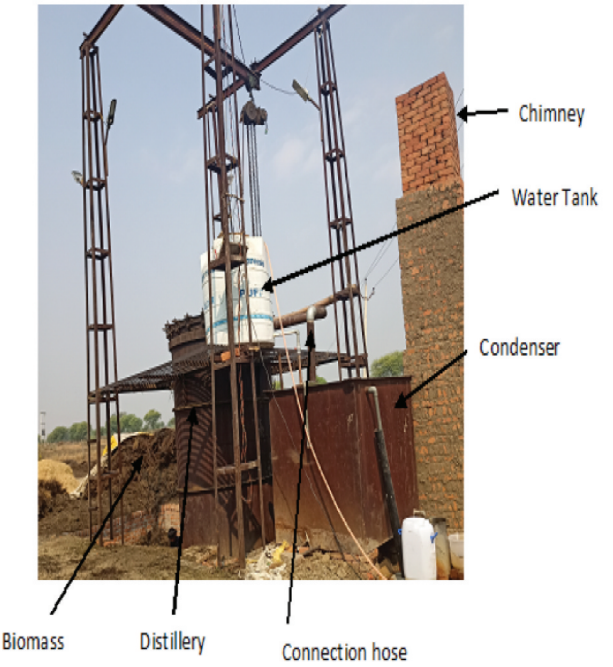


Figure 3. Experimental setup of conventional distillation system

Table 2. Description of conventional distillation system

S. No.	Name of component	Material	Size (meters)	Working description
1	Furnace	Bricks and concrete	1.82 × 2.27 × 0.076	Heat is generated by burning biomass (wood) in the furnace, transferred to the boiler.
2	Boiler	Iron	2.13 × 2.13 × 0.063	Boiler completes the evaporation process of water. Water in the boiler is heated up to the boiling point and converted into steam. Water required to complete one cycle was about 300–500 L. It is placed above the furnace.
3	Distillery	Stainless steel	Height = 3.04 Diameter = 1.82	Peppermint leaves are placed in two trays above the boiler in the distillery. One is kept at bottom and another at mid-height of distillery column. There are two openings in the distillery; one was at the bottom and the other at distillery. Opening at the bottom is used to pass the steam generated in the boiler to leaves of Peppermint, while top opening is used to pass the evaporated oil and water vapors to the condenser. The holding capacity of the distillery is about 1800 − 2000 kg batch size.
4	Condenser	Stainless steel	2.13 × 1.52 × 1.52	Mixture of oil and water vapors generated in the distillery is transferred to the condenser through the connection pipe. A water tank is attached to the condenser for continuous supply of water. Two outlets are used in a condenser to remove the hot water. Oil and water are separated in Florentine flask.

Observations

CSD system is installed at Birpura village, Jhansi, India (25° 36 N′18.40″E). Experiments are performed for one complete cycle during 3 days in November 2020 for various batch sizes of the entire plant (1500 kg, 1700 kg, and 1900 kg). Time duration of the experiments was the same for all three days (03:00 pm to 08:00 pm). Yield, mass of fuel, and temperatures at different positions in boiler (using thermocouples and digital temperature indicators), and steam temperature at distillery outlet are recorded during the experimentation. A beaker (measuring scale marked on upper side) and digital weighing machine measure the oil yield and mass of wood fuel, respectively. Dry branches of mango trees are used as fuel for burning.

Uncertainty analysis

Uncertainty is the most important factor in determining the accuracy of the experimental data. There are two types of data uncertainty: type A is random error, and type B is systematic error. Type B uncertainty is calculated here because data are uniformly distributed. The standard uncertainty of the instruments used for observation is calculated from Eq. (1). The accuracy and standard uncertainty of the instruments used are given in Table 3 (Kushwah et al. 2021).

$$StandardUncertainty = \frac{Accuracyofinstrument}{\sqrt{3}} \tag{1}$$

When any function Y depends on the number of input parameters or independent variables, then the uncertainty in the measured value of the function Y is given by Eq. (2) (Kushwah et al. 2021):

Table 3. Technical specification of the instruments used in the experiments

Instruments used	Accuracy	Range	Standard uncertainty
Thermocouple (TC-KT-02)	±0.1°C	−20 to 150°C	±0.24°C
Weight machine (HN-283)	0.1 gm	0.0001 to 20 kKg	0.00058

$$U(Y) = \sqrt{\left[\left(\frac{\partial Y}{\partial x_1}\right)^2 u^2(x_1) + \left(\frac{\partial Y}{\partial x_2}\right)^2 u^2(x_2) + \ldots\ldots + \left(\frac{\partial Y}{\partial x_n}\right)^2 u^2(x_n)\right]} \quad (2)$$

where U(Y) is the uncertainty in the measured function and $u(x_1) \ldots\ldots u(x_n)$ is the uncertainties in the independent variables ($x_1 \ldots\ldots x_n$), affecting the function Y.

Response surface methodology (RSM)

RSM was utilized to optimize process parameters that affect the EOY. Two process parameters were selected: extraction time A (60, 180, and 300 min) and batch size B (1500,1700, and 1900 kg) to obtain the maximum EOY from peppermint leaves. Prior experiments determined the range for every independent variable. Every parameter was examined at three levels: lower (−1), higher (+1), and central values (0), as shown in Table 4. The experimental data were designed using a design expert software 11.0 with face-centered central composite design (CCD). This created 13 trials to determine the pure error. The extraction yield (%) was chosen as the response (dependent variable). Table 4 shows the independent variables considered for optimization. Table 5 shows the various operating parameters of the experiments conducted in CCD.

Steam distillation system performance

Performance analysis of CSD system is performed in this section. It is useful for energy-efficient operation. Following parameters are considered to evaluate the performance of CSD system.

Table 4. Experiments range and levels of input parameters for optimization

Extraction parameters	Units	Symbol	Coded levels		
	Lower level (−1)	Central value (0)	Higher Level (1)		
Extraction time	min	A	60	180	300
Batch sizes	kg	B	1500	1700	1900

Table 5. No. of runs of CCD

	Factor 1	Factor 2	Response 1
Run	A:Extraction Time	B:Batch Size	Yield
	min	kg	%
1	180 (0)	1700 (0)	0.79
2	60 (−1)	1700 (0)	0.1
3	180 (0)	1700 (0)	0.8
4	300 (1)	1700 (0)	1.06
5	300 (1)	1500 (−1)	0.89
6	300 (1)	1900 (1)	1.05
7	180 (0)	1500 (−1)	0.66
8	180 (0)	1700 (0)	0.81
9	60 (−1)	1900 (1)	0.2
10	180 (0)	1700 (0)	0.78
11	60 (−1)	1500 (−1)	0
12	180 (0)	1700 (0)	0.82
13	180 (0)	1900 (1)	0.78

Thermal efficiency of conventional distillation system

Efficiency is the ratio of output to input. The yield of peppermint oil is the output and heat energy of wood fuel is the input in this case. Thermal efficiency of the system is evaluated as Eq. (3) (Dejchanchaiwang et al. 2019):

$$\eta = \frac{\dot{m} \times L}{m_f \times C.V.} \times 100 \tag{3}$$

Operational performance of conventional distillation system

System productivity, EOY, and extraction efficiency are estimated to conduct the operational performance of the conventional distillation system.

System productivity

Total amount of oil extracted from peppermint leaves is the system productivity. Hourly system productivity is evaluated from Eq. (4) (Radwan et al. 2020a, 2020b).

$$P = \frac{V}{t} \tag{4}$$

Cumulative system productivity is evaluated by recording the yield (L) at the end of the extraction stage.

Essential oil yield

EOY of the system is defined as the ratio of mass of extracted oil to mass of the plant per batch size and determined by the following Eq. (5) (Radwan et al. 2020a, 2020b; Munir et al. 2014).

$$EOY = \frac{m_{oil}}{m_{plant}} \tag{5}$$

In Eq. (5), m_{oil} is the mass of oil in kg and m_{plant} represents a mass of plant per batch size in kg.

Extraction efficiency

Extraction efficiency is the ratio of oil mass to maximum mass of oil per plant batch size. Efficiency of oil extraction is evaluated by 100 g of peppermint carries 1 mL of volatile oil and calculated from Eq. (6) as (Radwan et al. 2020a, 2020b):

$$\eta_{extraction} = \frac{m_{oil}}{m_{\max oil}} \tag{6}$$

In Eq. (6), $m_{\max oil}$ represents the maximum mass of oil per plant batch size in kg.

Environmental impact

Energy consumption and carbon dioxide (CO_2) emissions are determined to assess the environmental impact. The energy consumption for CSD method was calculated as (Drinic et al. 2020):

$$E_c = E \times t \tag{7}$$

$$E = m_f \times C.V. \tag{8}$$

According to Drinic et al. (2020), 800 g of CO_2 is rejected to the environment during burning coal or fossil fuel to produce 1 kWh. CO_2 emissions can be calculated by the following equation (Drinic et al. 2020):

$$E_{CO_2} = (E_c \times 800)/1000 \tag{9}$$

Results and discussion

Thermal performance of conventional distillation system

Thermal efficiency of the conventional distillation system is evaluated from Eq. (3). Hourly thermal efficiency for various batch sizes of peppermint plants for three different days is illustrated in Figure 4.

Thermal efficiency increases with the increase in plant batch size (Figure 4). It varies from 0 to 1.214%, 0 to from 1.804%, and from 0 to 2.081% for 1500 kg, 1700 kg, and 1900 kg batch size. Trend for all curves is similar for all batch sizes. Maximum thermal efficiency was 2.0% for 1900 kg batch size at 7–8 pm. Hence, it increases due to an increase in the system's yield and reduced fuel consumption per hour.

Operational performance of conventional distillation system

Operational performance of the conventional distillation system is evaluated by measuring the system's productivity, EOY, and extraction efficiency. Hourly productivity of the unit at different batch sizes of peppermint plants for three different days is illustrated in Figure 5.

Figure 5 indicates that the maximum hourly productivity of the system is 8.3 L/h at the time interval of 4–5 pm for the batch size of 1900 kg. It increases for the first 2-h interval in the initial then decreases simultaneously. Hourly system productivity increases from 5.5 to 8.3 L/h with the increase in batch size from 1500 kg to 1900 kg during 4–5 pm. Hourly system

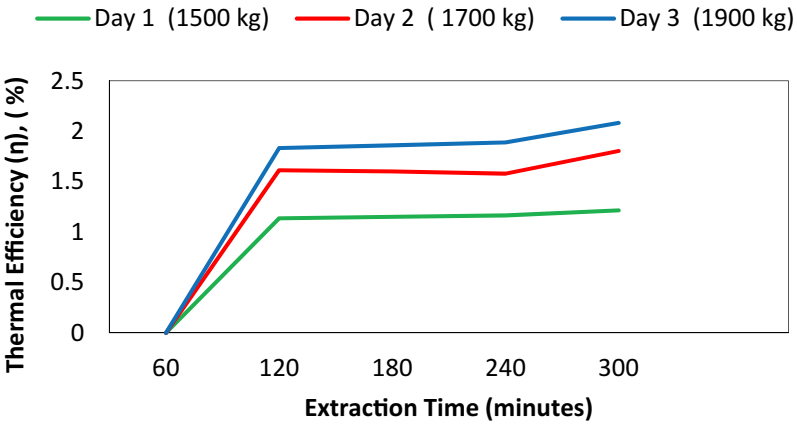


Figure 4. Variation in thermal efficiency with extraction time for different batch size

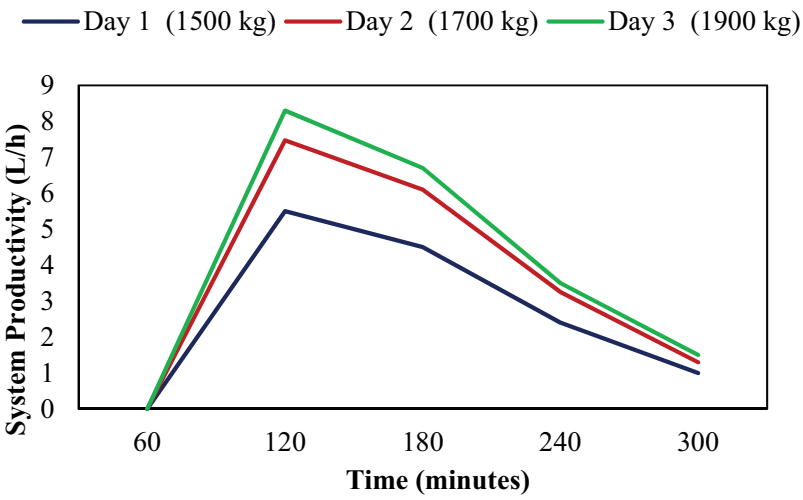


Figure 5. Hourly productivity of system with extraction time at various batch sizes

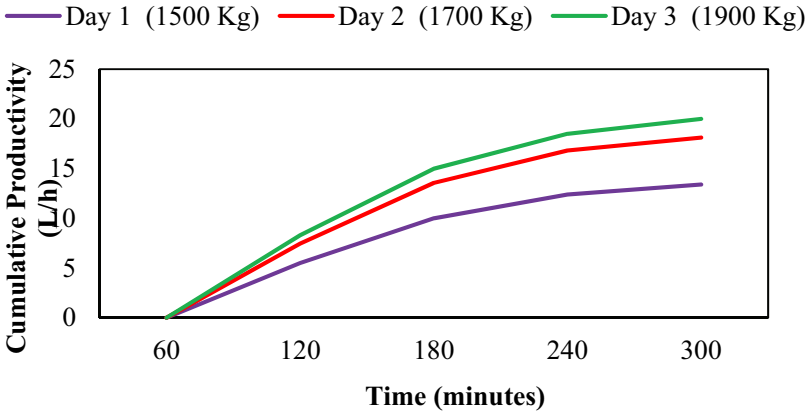


Figure 6. Cumulative productivity of system with extraction time at different batch sizes

productivity increased with the increase in plant batch size because generated steam increased in a distillation still. This generated steam accelerated the steam motion within the distillery, due to which volatile oil is extracted easily.

Cumulative productivity of the system at different batch sizes of plants for different time intervals is shown in Figure 6. It increases simultaneously with an increase in time and batch sizes of plant. Maximum cumulative productivity of the system is 20 L/h for 1900 kg plant batch size. It increases from 13.4 L/h to 20 L/h with an increase in batch size from 1500 kg to 1900 kg during 7–8 pm. Extraction efficiency of the system for different batch sizes is shown in Figure 7.

Experimental results show that the extraction efficiency of the system is increasing with the increase in batch sizes of plants. It increases for the first 2 h intervals then decreases simultaneously with time. Maximum extraction efficiency is 41.5% for 1900 kg batch size at the time interval of 04:00–05:00 pm. Extraction efficiency increases from 41.04 to 41.5% for batch size

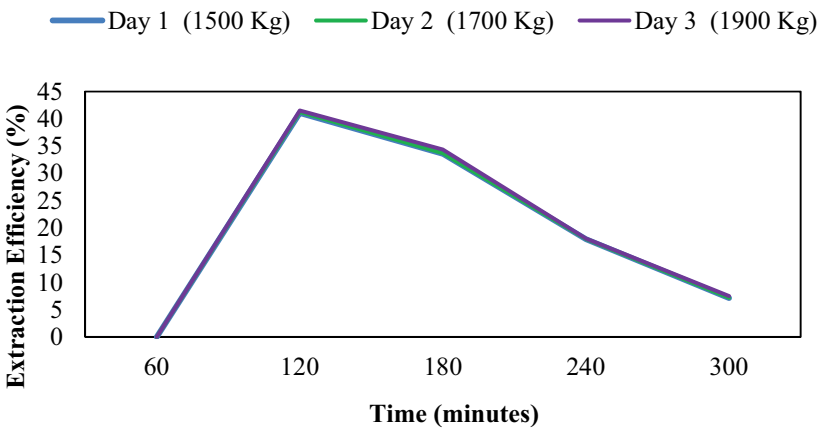


Figure 7. Variation in extraction efficiency of the system with extraction time

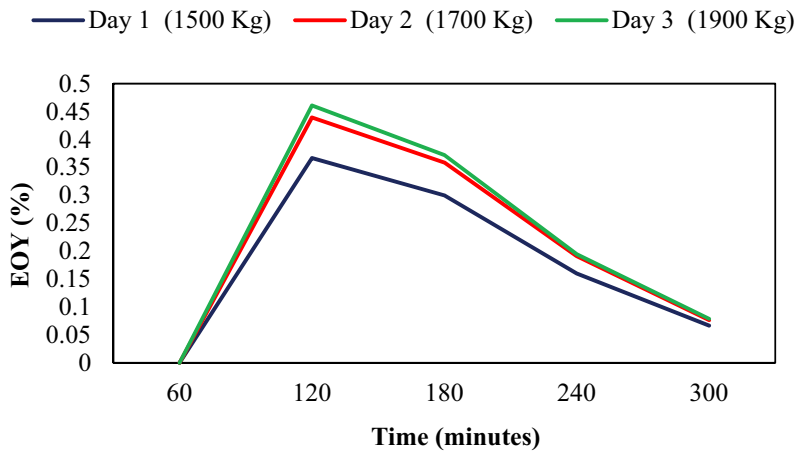


Figure 8. Variation in essential oil yield of system with extraction time

from 1500 kg to 1900 kg. There is a slight difference in the value of extraction efficiency due to a minor change in the batch size values. Therefore, the values of extraction efficiencies coincide in Figure 7. Average increase in extraction efficiency is 1.12%, with the increase in batch size from 1500 to 1900 kg at 04:00–05:00 pm.

Figure 8 illustrates the EOY of the system for different batch sizes of the plant at various time intervals. EOY increases with an increase in the batch size of the plant; however, all curves' trends are similar for all batches. Maximum EOY is 0.461% for the batch size of 1900 kg. EOY increases by 26% with the increase in batch size from 1500 to 1900 kg at 04:00–05:00 pm. EOY increases due to a steam rise generated, enabling volatile oil extraction by the rising batch size. Maximum EOY is obtained at 04:00–05:00 pm because maximum amounts of volatile oil are extracted at that time.

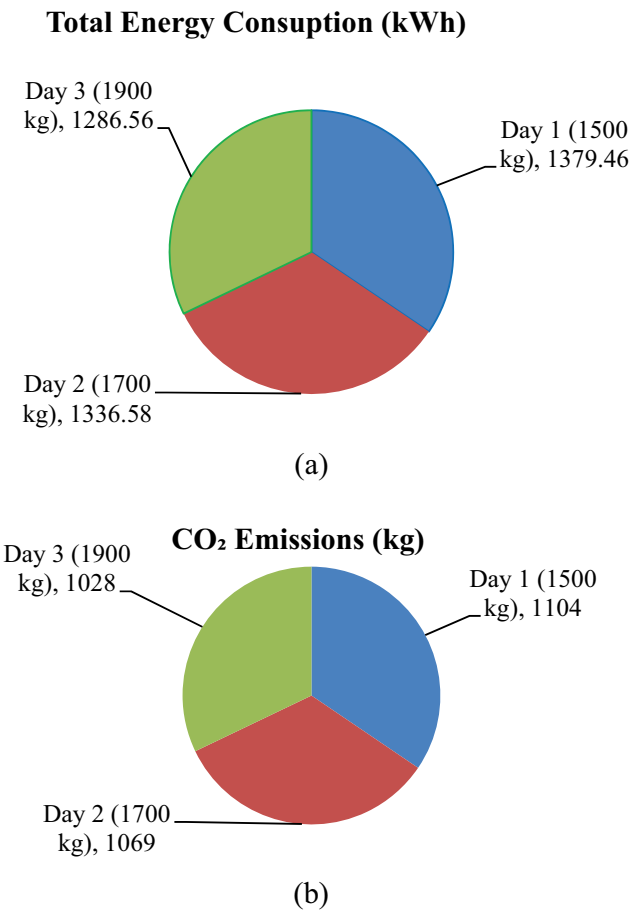


Figure 9. Environmental impact (a) Total energy consumption (b) CO_2 emissions

Table 6. Comparison of performance parameters of present study with earlier research

Productivity (mL)	Essential oil yield (%)	Extraction efficiency (%)	System efficiency (%)	References
6.2	0.797	88.57	-	[3]
7.3	0.785	98.13	60.25	[2]
28.20	-	-	33.21	[15]
-	0.40	-	-	[16]
1	3.7	-	-	[17]
39	-	-	-	[18]
20000	0.461	41.5	-	Present study

Environmental impact

Total energy consumption (TEC) and CO_2 emissions for one complete cycle of CSD system were calculated by Eqs. (7) and (9), respectively, to assess the environmental impact. Different batch sizes (1500, 1700, and 1900 kg) were used to analyze the environmental impact of SD system. TEC and CO_2 emissions for different batch sizes are depicted in Figure 9.

TEC and CO_2 emissions are decreasing with an increase in batch size, as shown in Figure 9. TEC is decreasing with the increase in batch size. Therefore, CO_2 emissions are also decreasing. TEC for 1500, 1700, and 1900 kg batch sizes are 1379.46, 1336.58, and 1286.56 kWh, while CO_2 emissions are 1104, 1069, and 1028 kg, respectively.

Overall performance of CSD system to extract the peppermint oil is compared with previous research work and found that EOY and extraction efficiency were 0.797% and 88.57% (Radwan et al. 2020a), 0.785% and 98.13% (Radwan et al. 2020b), 0.40% (Afzal et al. 2017), respectively. In the present study, EOY and extraction efficiencies were 0.461% and 41.5%. The comparison shows that results obtained in the present study are not much lower than those obtained in previous research on peppermint oil extraction. Performance parameters of present and previous studies are given in Table 6. Peppermint oil extraction units in India are still based on old technology. These systems need attention toward improving the design. The present study is performed at large-scale level and already published studies were performed at small-scale level. Hence, productivity is higher compared to previously published studies. Extraction efficiency and EOY of published work are higher than the present study because distillery was not insulated; therefore, losses are more. There is still a need to improve the design of peppermint oil extraction unit in Indian area.

Optimization using RSM

Response achieved for every run of CCD, coded experiments, influence of ET, and batch size on extraction yield were investigated. Table 5 represents experimental oil yield (%) under various permutations of extraction conditions. The highest and lowest EOYs were found in experiment number 4 and 11, respectively. Table 5 summarizes the data obtained after experimental design.

Table 7. ANOVA for polynomial model

Source	Sum of squares	df	Mean square	F-value	p-value	
Model	1.43	5	0.2863	348.85	< 0.0001	Significant
A-extraction time	1.21	1	1.21	1480.46	< 0.0001	
B-batch size	0.0384	1	0.0384	46.79	0.0002	
AB	0.0004	1	0.0004	0.4874	0.5076	
A^2	0.1139	1	0.1139	138.82	< 0.0001	
B^2	0.0110	1	0.0110	13.40	0.0081	
Residual	0.0057	7	0.0008	-	-	
Lack of fit	0.0047	3	0.0016	6.33	0.0534	Not significant
Pure error	0.0010	4	0.0002	-	-	
Cor total	1.44	12				

Table 8. Fit statistic.

Std. dev.	0.0286	**R^2**	0.9960
Mean	0.6723	**Adjusted R^2**	0.9931
C.V. %	4.26	**Predicted R^2**	0.9673
		Adeq precision	54.4644

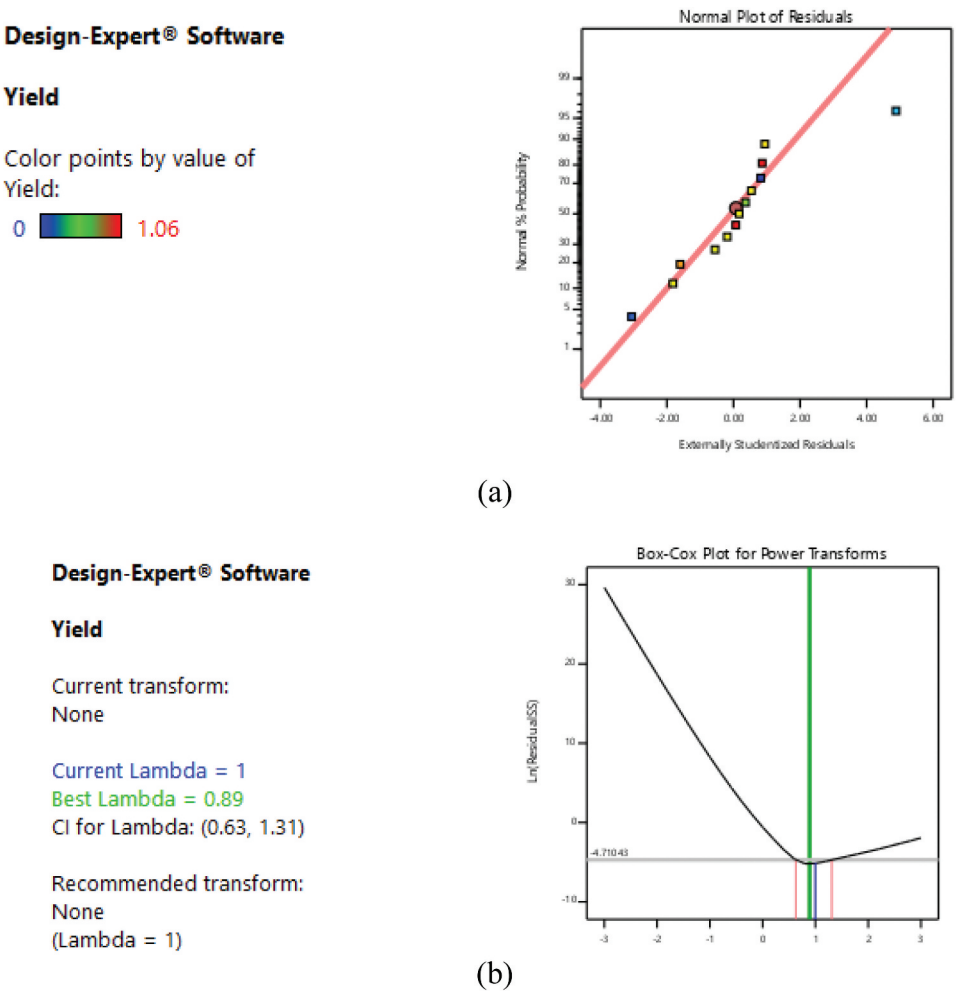


Figure.10. (a) Normal plot of residuals and (b) Box-Cox plot for power transform

Fitting the model

RSM was utilized to obtain the optimal condition of process parameters to achieve the maximum peppermint oil yield. A quadratic polynomial equation was used for fitting the observed data of analyzed variables. The response surface analysis used a generalized second-order polynomial model, which is written as:

$$Y = 0.7952 + 0.45A + 0.08B - 0.01AB - 0.2031A - 0.0631B \tag{10}$$

where Y is the extraction yield (response) and A and B are the independent variables. The number of CCD runs performed in this study is shown in Table 5.

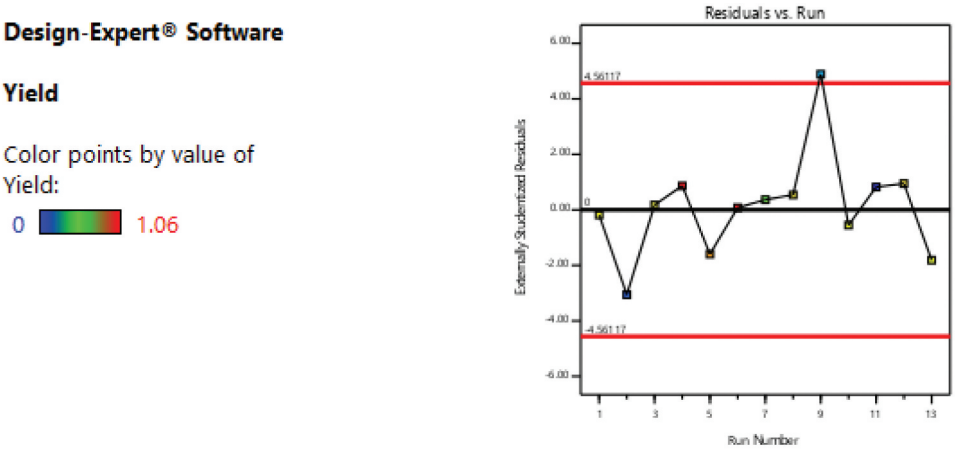


Figure 11. Studentized residuals vs. (a) run number and (b) predicted response for yield of peppermint oil

ANOVA for the model

Table 7 and 8 shows the ANOVA of the resultant polynomial model of the observational data. P-values ($P<0.05$) indicate that model is significant. Model F-value is 22.56, indicating that the model is statistically significant and there is only a 0.01 % chance that F- Value could occur due to noise. The Terms A, B, A^2 and B^2 are significant ($P<0.1$) in this situation. The model term AB is not relevant ($P>0.1$). The lack of Fit F-value of 6.3 indicates that the lack of fit is not considerable. Results of present study indicate that coefficient of determination (R^2) was 0.9960 for the regressive model. This indicates that independent variables are responsible for 99.56% of sample variation and only 0.44% of total variation does not explain by this model. A high R^2 shows that the data can adequately account for the variation to fit the model. The predicted model appeared to be a good fit for the observed data. As a result, the model adequately explained the response. Model's accuracy is not implied by considering only R^2 because adding a variable in model increases the R^2. Therefore, Adj-R^2 is better to evaluate the model if R^2 value is greater than 90% (Belhachat et al. 2018). In this study, both R^2 (0.9960) and Adjusted R^2 (0.9931) indicated that quadartic regression model suited the response values well and

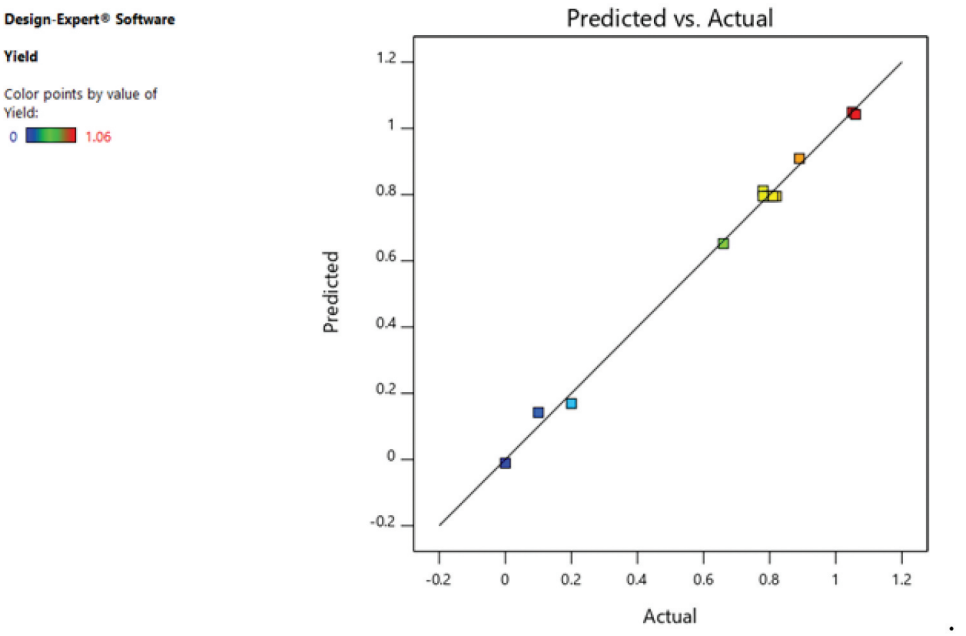


Figure 12. Actual and predicted values for yield

that the observed and predicted values are consistent. Model is acceptable since the predicted R^2 of 0.9673 is in reasonable agreement with the Adjusted R^2 of 0.9931 i.e. the difference is less than 0.2 (Table.8). Signal to noise ratio is measured by "Adeq Precision". "Adeq Precision" of 54.464 (>4) implies that model is acceptable and used to explore the design space. Coefficient of variation (CV)<10 suggested that the model is reproducable (Belhachat et al. 2018). CV of 4.26 values implies that experimental data is reliable and model is repeatable.

Studentized residuals vs. probability curve is shown in Figure 10(a). This figure illustrates that the data are spread over a line and indicate that residuals follow a normal distribution. Figure 10(b) illustrates the Box-Cox curve for power transform. The plot shows best lambda 0.89 with lower and higher confidence intervals as 0.61 and 1.31, respectively. Distribution of studentized residuals against run number and predicted yield of peppermint oil is illustrated in Figure 11. Studentized residuals follow a random pattern with respect to run numbers.

Figure 12 represents the relationship between actual and predicted values for peppermint oil yield. Predicted values are scattered close to the 45° line, indicating the polynomial model's good predictive accuracy. The actual and predicted yield values lie equally on both sides of the line.

Contour and response surface plots

Response surface and contour plots clearly illustrate the relationship between dependent and independent variables. 2D contour plots and 3D response surface plots are used to depict the interactive effects of process parameters on extraction yield. Furthermore, the plots are used to estimate the maximum yield. One parameter remains constant at central value while the other two vary (Phat et al. 2020). In this study, the combined effect of ET and batch size on extraction yield of peppermint oil is shown in Figures 13(a) and 13(b). This study indicates that extraction yield increases with the increase in batch size and ET, but increasing ET shows

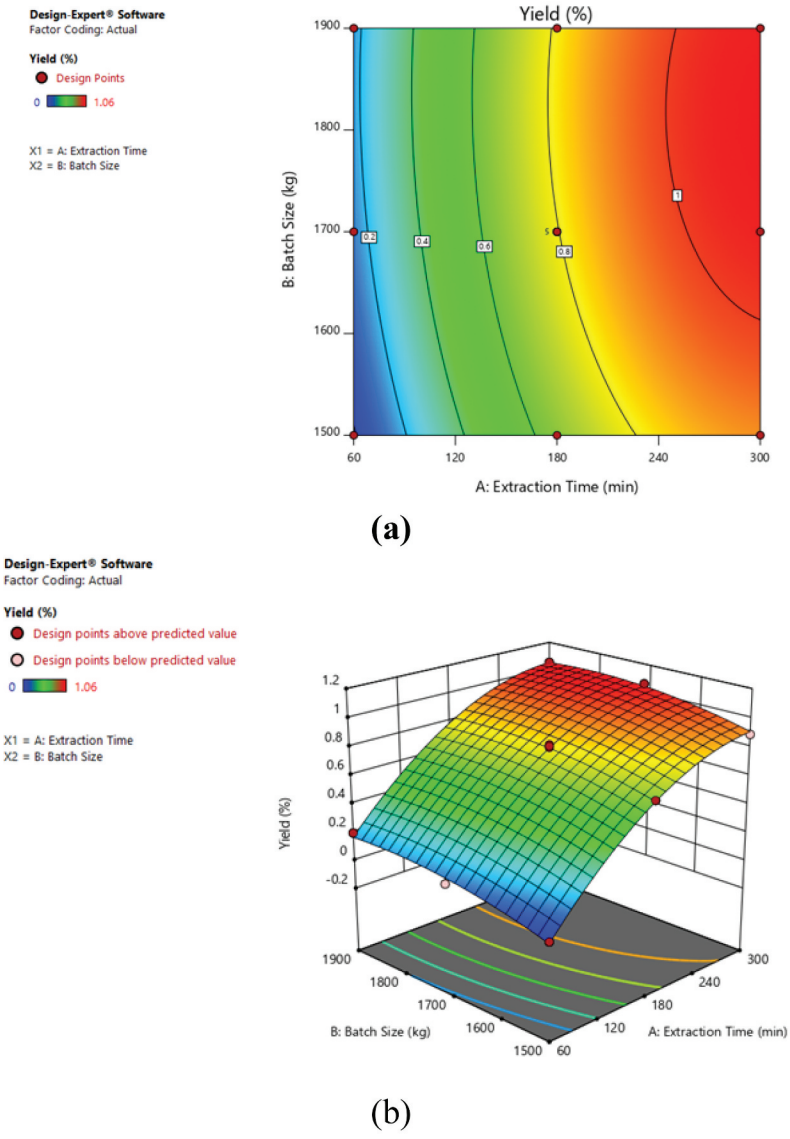


Figure 13. Contour and response surface plot: (a) 2D and (b) 3D

a larger effect on yield than batch size. The contour plot shows that the oil yield could reach approximately 1.06% when ET ranges from 238 to 300 min and batch size from 1600 to 1900 kg.

Optimal condition

Out of 34 solutions generated by software, one is selected as an optimal solution at maximum oil yield. Red-colored points show the optimal conditions (ET and batch size) and the maximum peppermint oil yield shown by blue colored dot in Figure 14(a). Desirability of solution is 1 as shown in Figure 14(b). This indicates that these conditions are the best optimal solutions for present study. As per the results of RSM, the maximum yield of 1.06% was obtained at 300 min ET and 1807.5 kg batch size of peppermint plants. Experiments were

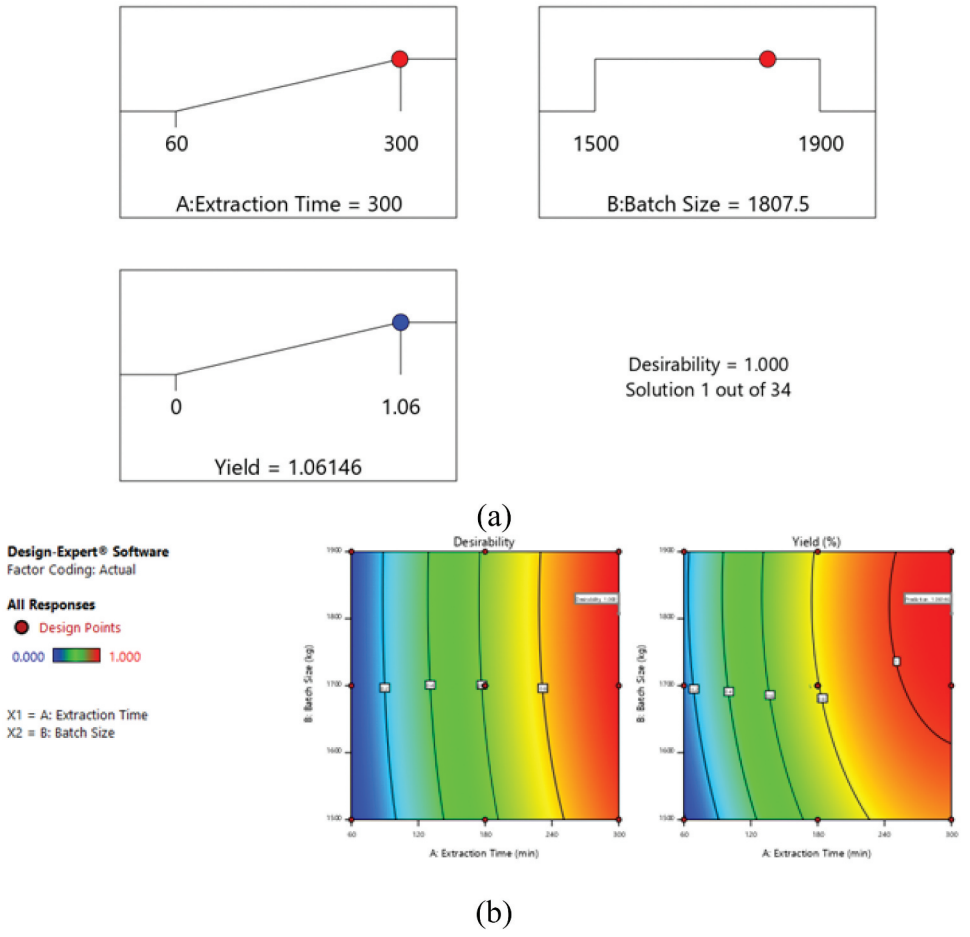


Figure.14. (a) Optimal conditions and (b) contour plot of desirability of solutions

further performed for 1807.5 kg and the yield was found as 1.04621%. Predicted value of yield was obtained as 1.06146%. Experimental and predicted yield values are in good consistency with percentage deviation of 1.436.

Conclusions

Following conclusions have been drawn based on the experimental results and analysis:

- System productivity (20 L), maximum hourly thermal efficiency (2%), extraction efficiency (41.5%), and maximum EOY (0.461% (w/w)) were obtained for a batch size of 1900 kg.
- Thermal efficiency, cumulative system productivity, EOY, and extraction efficiency are increased with the increase in plant batch size.
- Average increase in total productivity, extraction efficiency, and EOY was 49.25%, 1%, and 26%, with an increase in batch size from 1500 kg to 1900 kg.
- TEC for 1500, 1700, and 1900 kg batch sizes are 1379.46, 1336.58, and 1286.56 kWh, while CO_2 emissions are 1104, 1069, and 1028 kg, respectively.
- TEC and CO_2 emissions are decreasing with an increase in batch size.

- A quadratic polynomial model was developed for predicting the extraction yield and showed a good agreement between actual and predicted values of extraction yield (R^2=0.9960).
- As per ANOVA, the impacts of extraction time, batch size, and their interactions were the key process parameters shown to be statistically significant in influencing the EOY.
- Maximum oil yield of 1.06% was obtained at optimal parameters at 300 min ET and 1807.5 kg batch size.

Therefore, based on the present study's conclusions, there is an urgent and immediate need to improve the design of Indian distillery units used for peppermint extracting oil to enhance their performance. Implementation of solar thermal energy is also recommended, along with peppermint biomass waste generated during the Indian distillery unit process, to make energy-efficient, sustainable, and environment-friendly SD systems.

Nomenclature

$\dot{m}$	Yield (kg/h)
L	Latent heat of vaporization of Peppermint (kJ/kg)
m	Mass (kg)
C.V.	Calorific value of wood fuel
P	Productivity of the system
V	Volume of oil (L)
t	Time (h)
f	Wood fuel
EOY	Essential oil Yield
CSD	Conventional steam distillation
TEC	Total energy consumption
ET	Extraction Time
PO	Peppermint oil
EO	Essential oil
SSR	Solid to solvent ratio
UP	Ultrasonic power
ANOVA	Analysis of variance
η	efficiency

Acknowledgments

The first author would like to thank Delhi Technological University, Delhi, for awarding the fellowship for pursuing the Ph.D. in the Mechanical Engineering Department and gratefully acknowledge Centre for Energy and Environment to support the present work.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The authors have no funding to report.

ORCID

Ravi Kant http://orcid.org/0000-0001-5494-0725
Anil Kumar http://orcid.org/0000-0002-2846-4347

Availability of data and material (data transparency)

This is a review article of previous work. Critical analysis has been done with reference to earlier research work. Hence, there is no data used.

Authors' contributions (optional)

Ravi Kant: Methodology, Writing - original draft, Investigation.
Anil Kumar: Conceptualization, Visualization, Writing - review & editing, Supervision, Investigation.

References

Afzal, A., A. Munir, A. Ghafoor, and J. L. Alvarado. 2017. Development of hybrid solar distillation system for essential oil extraction. *Renewable Energy* 113 (Dec):22–29. doi:10.1016/j.renene.2017.05.027.

Belhachat, D., L. Mekimene, M. Belhachat, A. Ferradji, and F. Aid. 2018. Application of response surface methodology to optimize the extraction of essential oil from ripe berries of Pistacia lentiscus using ultrasonic pretreatment. *J. Appl. Res. Med. Aromat. Plants* 9 (November):132–40. doi:10.1016/j.jarmap.2018.04.003.

Dejchanchaiwong, R., A. Kumar, and P. Tekasakul. 2019. Performance and economic analysis of natural convection based rubber smoking room for rubber cooperatives in Thailand. *Renewable Energy* 132:233–42. doi:10.1016/j.renene.2018.07.145.

Drinić, Z., D. Pljevljakušić, J. Živković, D. Bigović, and K. Šavikin. 2020. Microwave-assisted extraction of O. vulgare L. spp. hirtum essential oil: Comparison with conventional hydro-distillation. *Food Bioprod. Process* 120 (2006):158–65. doi:10.1016/j.fbp.2020.01.011.

Fadhil, A. B., L. A. Saleh, and D. H. Altamer. 2020. Production of biodiesel from non-edible oil, wild mustard (Brassica Juncea L.) seed oil through cleaner routes. *Energy Sources, Part A Recover. Util. Environ. Eff* 42 (15):1831–43. doi:10.1080/15567036.2019.1604893.

Gavahian, M., and Y. Chu. 2018. Ohmic accelerated steam distillation of essential oil from lavender in comparison with conventional steam distillation. *Innov. Food Sci. Emerg. Technol* 50 (September):34–41. doi:10.1016/j.ifset.2018.10.006.

Işcan, G., N. Kirimer, M. Kürkcüoğlu, K. H. C. Başer, and F. Demirci. 2002. Antimicrobial screening of Mentha piperita essential oils. *Journal of Agricultural and Food Chemistry* 50 (14):3943–46. doi:10.1021/jf011476k.

Kant, R., and A. Kumar. 2021. Advancements in steam distillation system for oil extraction from peppermint leaves. *Mater. Today Proc* 47 (17):5794–99. doi:10.1016/j.matpr.2021.04.123.

Kubátová, A., D. J. Miller, and S. B. Hawthorne. 2001. Comparison of subcritical water and organic solvents for extracting kava lactones from kava root. *Journal of Chromatography A* 923 (1–2):187–94. doi:10.1016/S0021-9673(01)00979-7.

Kulturel, Y., and S. Tarhan. November 2016. Performance of a solar distillery of essential oils with compound parabolic solar collectors. *Journal of Scientific and Industrial Research* 75:691–96.

Kushwah, A., A. Kumar, M. K. Gaur, and A. Pal. 2021. Garlic dehydration inside heat exchanger-evacuated tube assisted drying system: Thermal performance, drying kinetic and color index. *Journal of Stored Products Research* 93 (May):101852. doi:10.1016/j.jspr.2021.101852.

Liu, X., H. Ou, Z. Xiang, and H. Gregersen. 2019. Optimization, chemical constituents and bioactivity of essential oil from Iberis amara seeds extracted by ultrasound-assisted hydro-distillation compared to conventional techniques. *J. Appl. Res. Med. Aromat. Plants* 13 (83). doi: 10.1016/j.jarmap.2019.100204.

Loolaie, M., N. Moasefi, H. Rasouli, and H. Adibi. 2017. Peppermint and its functionality : A review. *Arch. Clin. Microbiol* 8:1–16. doi:10.4172/1989-8436.100053.

Memarzadeh, S. M., A. Gholami, A. G. Pirbalouti, and S. Masoum. 2020. Bakhtiari savory (Satureja bachtiarica Bunge.) essential oil and its chemical profile, antioxidant activities, and leaf micromorphology under green and conventional extraction techniques. *Industrial Crops and Products* 154 (January):112719. doi:10.1016/j.indcrop.2020.112719.

Munir, A., O. Hensel, W. Scheffler, H. Hoedt, W. Amjad, and A. Ghafoor. 2014. Design, development and experimental results of a solar distillery for the essential oils extraction from medicinal and aromatic plants. *Sol. Energy* 108:548–59. doi:10.1016/j.solener.2014.07.028.

Peterson, A., S. Machmudah, B. C. Roy, M. Goto, M. Sasaki, and T. Hirose. 2006. Extraction of essential oil from geranium (Pelargonium graveolens) with supercritical carbon dioxide. *Journal of Chemical Technology & Biotechnology* 81 (2):167–72. doi:10.1002/jctb.1375.

Phat, D. T., K. C. Tuyen, X. P. Huynh, and T. T. Truc. 2020. Extraction process optimization and characterization of the pomelo (Citrus grandis l.) peel essential oils grown in tien giang province, Vietnam. *Nat. Volatiles Essent. Oils* 7 (4):26–33. doi:10.37929/nveo.780505.

Radwan, M. N., M. M. Morad, M. M. Ali, and K. I. Wasfy. 2020a. A solar steam distillation system for extracting lavender volatile oil. *Energy Reports* 6:3080–87. doi:10.1016/j.egyr.2020.11.034.

Radwan, M. N., M. M. Morad, M. M. Ali, and K. I. Wasfy. 2020b. Extraction of peppermint volatile using a sample steam distillation system. *Plant Arch* 20 (i):1487–91.

Rajeshwaran, M., R. Parthasarathi, J. Ganeshkumar, and A. Ponshanmugakumar. 2020. A multi-variant approach to optimize process parameters for oil extraction from non edible feedstock: Acacia nilotica. *Materials Today: Proceedings* 33:4671–76. doi:10.1016/j.matpr.2020.08.314.

Riachi, L. G., and C. A. B. De Maria. 2015. Peppermint antioxidants revisited. *Food Chemistry* 176 (Jun):72–81. doi:10.1016/j.foodchem.2014.12.028.

Sekhar, S. C., K. Karuppasamy, M. V. Kumar, D. Bijulal, N. Vedaraman, and R. Sathyamurthy. 2021. Rain tree (Samanea saman) seed oil: Solvent extraction, optimization and characterization. *J. Bioresour. Bioprod* 0–16. doi:10.1016/j.jobab.2021.04.005.

Sodeifian, G., N. Saadati Ardestani, and S. A. Sajadian. 2019. Extraction of seed oil from Diospyros lotus optimized using response surface methodology. *Journal of Forestry Research* 30 (2):709–19. doi:10.1007/s11676-018-0631-8.

Sodeifian, G., S. A. Sajadian, and N. Saadati Ardestani. 2017. Experimental optimization and mathematical modeling of the supercritical fluid extraction of essential oil from eryngium billardieri: Application of simulated annealing (SA) algorithm. *J. Supercrit. Fluids* 127 (January):146–57. doi:10.1016/j.supflu.2017.04.007.

Tran, T. H., H.H.H. Nguyen, D.C. Nguyen, T.Q. Nguyen, H. Tan, L.T.H. Nhan, D.H. Nguyen, L.D. Tran, S.T. Do, T.D. Nguyen. 2018. Optimization of microwave-assisted extraction of essential oil from Vietnamese basil (Ocimum basilicum L.) using response surface methodology. *Processes* 6 (11):1–12. doi:10.3390/pr6110206.

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Ranking Users in the Social Network: A Survey of Budget and Opinion dependent Methods

Akshit Tomar
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
akshittomar3002@gmail.com

Aman Dhiman
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
amandhiman262000@gmail.com

Abhishek Midha
Dept. of Computer Science and Engineering
Delhi Technological University
Delhi, India
abhi.midha777@gmail.com

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760915

***Abstract*—Social networks have become a popular abstraction for capturing individual user interactions in various everyday activities and applications. As a corollary, network analysis has gained a considerable attention in the scientific community. Identifying influential users for a website is one of the subjects of interest. A variety of uses, including message distribution, viral marketing, or preventing the spread of rumours. Several options have been suggested. To determine the influence of users and to discover groups of prominent people in social media networks. Although several surveys have examined ways based on social network,structural aspects, there has been no systematic examination of approaches that consider user behaviour. This study intends to fill the gaps by evaluating and proposing a taxonomy of such behaviour-cognizant approaches for identifying prominent members in social networks.**

***Index Terms*—Social Networks, Influence Maximisation,Budget Dependent,Opinion dependent,Trust dependent**

I. INTRODUCTION

People's daily routines and patterns of engagement have changed due to the growth of network technology, particularly social media. Various types of online social networks have attracted many users who regularly share large amounts of data. Social networks can be used for a variety of reasons, including exchanging messages, disseminating news, expressing ideas on topics of mutual interest, publicizing events, link prediction, and ranking users [1, 2]. Because of some users' behaviour or network of friends, they are more famous have an influential say on a particular topic. In social networks, so-called powerful individuals play a significant role in helping to propagate messages fast and broadly. The difficulty of selecting a collection of users in a social network who may influence other users extensively and effectively are ranked on in the social network. This is known to be a complicated subject, mainly because it necessitates a criterion for determining influence. Peoples' power can be measured using a variety of parameters. In some cases, According to studies, the network structure determines the user's effect, which means essential users are recognized based on different configurations of the communication network. A lot of these surveys show unique insights into independent behaviour methods. The general truth is that argumentative behaviours have a long past in tackling the Influence Maximization problem. These behaviour-dependent techniques possess the potential to provide significant answers to real-life situations. These plans haven't been looked into in past surveys as their own category. Our study attempts to fill that need.

II. PREFATORY

A. Problem Description

Relationships between different users on a social network show a lot of different parameters: such as interests, ties, bonds, and attachment between multiple users. With the knowledge that the social network provides us with, we can rank users based on their influence and top-ranking individuals may be used as an initial seed for spreading of information and influencing the masses.The initial seed seizes the message and causes it to spread. A communication network graph is modelled as $G = (V, E, FV, RE)$, where users and different links between them are defined as $V = v_1, v_2, ..., v_n$ and edges $E = e_1, e_2, ..., e_n$ respectively.$Fv_{|V|} = Fv_1, Fv_2, ..., Fv_{|V|}$ is a set of various vectors where $Fv_i = fv_1^{(i)}$,$fv_2^{(i)}$,..., $fv_{|V|}^{(i)}$ represents different features of nodes in the network , which may include geographical coordinates, likes ,dislikes,beliefs and more.$RE = Re_1, Re_2,...,Re_{1|E|}$ is another set of vectors where $Re_1 = re_1^{(ij)}$, $re_1^{(ij)}$,..., $re_{|R|}^{(ij)}$ represents the different relationships of the links $e_{(ij)}$ in the social network represented ,these include spreading probability,activation probability,trust and likes and dislikes between different dyad's .Some research considers the links to be directed, which means $e_{ij} \neq e_{ji}$, while the other research mostly considers the links to be conjoint, which makes the graph to be undirected, which means $e_{ij} = e_{ji}$. Ranking individuals in the network aim to find a small group of individual nodes which will be part of the initial seed set and will be responsible for the dissemination of a query in a network.The message is represented as $M = (Q, QF, B)$. Q represents the query message which may include various topics, such as advertising, news articles and much more$QF = qf_1, qf_2, ...qf_N$ includes various features of the query message. B is the budget constraint that restricts the seed's trigger ability.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

B. Ranking users

Ranking users in the social network is defined in the equation, $f(S)$ defines the spreading ability of collections of users U, usually the highest-ranking individuals in the network.

$$U^* = argmax_{U \subset V} f(U) \quad (1)$$

subject to some constraints

f is monotone if $f(U) \leq f(U)$ for all $U \subseteq T$.f is sub-modular if $f(U \cup v_i) - f(U) \geq f(T \cup v_i) f(T)$ for all $S \subseteq T$.

Many techniques are written in this paper to Rank the influential individuals in the networking model and to find out the top-ranking individuals who can be used as seeds. We divide these techniques into two different groups:

- **Centrality measure** determines the value of each node on a public graph based on topological structures. Different steps of centrality measure are proposed and expanded to determine the impact of nodes and see the seed. Betweenness [10], entropy-based [43], h-index [12], ncvoterank [44], weighted voterank [45] are popular measures that can be used. K-shell [26] is another well-known measure that determines the average of the nodes based on the distance to the core of the graph. A number of intermediate steps have been proposed based on the k-shell, which includes the k-shell iteration factor, Modified-degree with mutual exclusion ratio (MDER), a mixed degree decay [25].
- Imitating the distribution process can be used to determine the impact in each node and select a set of influential notes. Every element inside this approach will either be activated or dormant. For every component v_i, a cutoff limitation is evaluated, as well as a load w_{ij} is applied to every path e_{ij}. At time zero $(t = 0)$, the seeds are alive, and other such components are dormant. When summation$\sum_{v_i \in BN_{ij}} w_{ij} > \theta_j$ at every moment t, every dormant component v_j becomes activated, whereas BN_j signifies the neighbours of component v_j who are alive in $time(t) = 1$. The procedure is performed recursively, and in every cycle, a user is introduced to U as a new recruit whose conjunction with U maximises the distributing capability, that is, the user with the largest marginal impact. In the $t-th$ cycle, Eq. (2) is used to discover the user with the highest moderating effect, e_m, where U_{t-1} is the seed-node set chosen in iteration $t-1$.

$$v_m = argmax_{v_i \in V - U_{t-1}} f(U_{t-1} \cup v_i) - f(U_{t-1}) \quad (2)$$

C. Proposed methods for Ranking users

The study divides Ranking individuals in a network into two groups: (a)behaviour-independent and (b)behaviour-dependent methods. Behaviour-independent algorithms simply analyse the graphs and figure out the effects of different user nodes and single out important nodes and form a group. Behaviour-dependent algorithms include various characteristics of user opinions and characteristics to suggest more realistic alternatives for implementation in the real world. In this work, our main focus will be behaviour-dependent approaches. Returning to our problem description, behaviour impartial techniques do not take into consideration anything special to the sets FV (recording behaviour of different users) and (RE) (grasping various characteristics of the links between a dyad). Furthermore, they do not take into account distinct aspects of the query, as stated in the set QF of the message. Finally, a seed set is generally supposed to have a certain fixed amount of entities while, in the model query, budget,B, it may also be related to various behavioural-partial factors.

III. Behaviour Dependent Techniques

Under behaviour-dependent approaches, additional data, which includes user interactions, connection traits, and inquiry data, is factored into the equation using $Fv_{|V|}$ and $Re_{|R|}$ additional to the underlying network to optimise the propagation phase. Whenever a message is designed to promote a new item, it's not like all consumers are increasingly keen on that item. Behaviour dependent methods are divided into two main categories: Opinion dependent and Budget dependent. The sentiment of users regarding a topic is put into consideration in opinion-dependent approaches. Budget-conscious strategies examine the expenses and advantages of disseminating a marketing question.

IV. Opinion-dependent Techniques

Among these techniques, every person has an impression or approach towards the inquiry. That is, FV_i has a value representing the user's views or attitudes. The sentiment is a restricted range of options and is frequently bipolar, denoted by a selection such as optimistic/pessimistic or a selection among viewpoints X/Y. A genuine value for an issue indicates perspective. In an election campaign, for example, users may have an inclination and choose contender X or contender Y. In this case, influence maximization can try to create a set of nodes that will cause signal propagation in such a manner that it maximizes the amount of people who will show support for contender X.It has frequently been stated that altering one's view or outlook is dependent on the person's trust. As a result, its been proposed to represent social media as a directed network graph with connections that seem to have a favourable or unfavourable sign indicating confidence or mistrust amongst users. However, not all solutions take trustworthiness into consideration. As a result, it is important to categorize opinion dependent approaches into two types: trust-dependent and trust-independent.

*1) **Trust-independent Techniques**:* In [34], FV_i displays user's behavioural data. FV_i has two configurations: the inner mentality that remains throughout time, and the exterior mentality, which is affected through time through other individuals. Every type of mentality is expressed by a numerical number in the range [0, 1]; a 0 indicates a wholly cynical outlook, while a score of 1 indicates a fully good mindset. The expense of altering every individual's mood is then calculated on the basis of both interior and exterior

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

mentality, and a greedy approach is suggested to locate impactful seed nodes. Every individual, as suggested by Feng et al. [36, 40] has an orientation toward the query specified by a numerical value in [1, +1]; furthermore, every link has a value reflecting the propagation likelihood across neighbouring individuals. Huiyuan et. al[40] proposes an opinion cascade model to mimic the propagation mechanism and how individuals' attitudes can change drastically. Whenever a site is engaged, Equation(3) is used to modify its mindset, where RE_{ij} denotes the value of the link connecting user v_j and v_i.

$$FV_j[B] = FV_j[B] + \sum_{v_i \in BN_j} RE_{ij}.FV_i[B] \tag{3}$$

Influence maximisation is represented as the challenge of locating a seeding set which, once engaged, maximises the quantity of enabled users with dispositions greater than a specified value. After that, the diffusion model is utilised to explain a greedy strategy for solving the issue. Wei Chen et al. [9, 32] suggests that every individual has a bad, unbiased, or good opinion regarding the query, i.e., $FV_i[O] \in [-1, 0, +1]$, and influence maximisation seeks to boost the efficiency with $+1$ opinions during the dissemination phase. The literature suggests that bad opinions are more powerful than good views in influencing user decisions. Therefore they expand the Independent Cascade diffusion model [23] and evaluate a seed-nodes set spreading capabilities. In IC-N[21], vertices inside a seed-nodes set are triggered, and their attitude is adjusted to $+1$ with the probability of p or -1 at probability $1 - p$; the remaining users are dormant and have attitude 0. The triggering mechanism is identical to that of the Independent Cascade diffusion model.

Every user can be triggered by a 0 value or $+1$ value node throughout dissemination. When a good node activates a user, its attitude transforms to a $+1$ with probability p or -1 with probability $1 - p$. When a bad node activates a user, its attitude transforms to -1. Ali et. al [32], contend that while distributing product advertisements, the function of consumers' criticising actions is ignored in the Negative IC model; a person could be happy with the item but still have few problems. As a result, they add a grievance element g to the Negative IC model.

*2) **Trust-dependent Techniques** :* Individuals may see other individuals as allies or adversaries, as described by Yanhua Li et al. [27]. Taking into account both pleasant and unpleasant connections might help to improve opinion-dependent approaches. In trust-dependent strategies, every person has their own set of beliefs regarding various queries, goods, and events. Certain views may gain or lose acceptance as a result of a person's friendship or enemy connection with other individuals. A symbol is often used to depict a friendly or enemy connection; links with a positive sign imply friendliness among users, whereas links with a minus sign suggest an enemy connection.As a result, $RE_{ij}[P]$ reflects the type of connection among nodes in addition to propagating probability. In other terms, links are seen as directional; every link e_{ji} from v_j to v_i demonstrates that v_i is impacted by v_j. RE_{ji} denotes the propagating likelihood from v_j to v_i, whereas $RE_{ij}[P]$ denotes faith or disdain from v_i to v_j. Huanhuan et al. [7, 31, 34] suggested that every individual has an opinion towards this query's information; that is, $FV_i[B] \in [1, +1]$ represents the individual's opinion towards to narrative. The Linear Threshold diffusion model [37] is extended in such studies to simulate transmission in a tagged network and to modify individuals' attitudes about dissemination. The nodes are triggered with outlook +1 at time(t) = 0 in this model; the triggering mechanism is similar to the classic Linear Threshold model. When a vertex v_j is active, however, its behaviour is modified as specified:

$$FV_j[A] = \sum_{v_i \in BN_j} FV_i.RE_{ij}[P].RE_{ij} \tag{4}$$

The ranking users problem's purpose is to maximise the quantity of seed nodes with $FV_i[B] > 0$. Hung et al. [34] surveyed both intrinsic $FV_i[IB]$ and expressive $FV_i[B]$ views for every individual v_i are considered, where $FV_i[IB]$ expresses inclination based on previous behaviour and does not alter during the propagating process. $FV_i[B]$ symbolises a mindset that may be affected and may change throughout dissemination as a consequence of the people's intrinsic mentality and the mentality of the proactive peers. In LT-S [37] Node s are triggered , and its stated mood is adjusted to +1 at time(t) = 0. When $time(t) > 0$, every inactive node v_j gets triggered in a method identical to the initial Linear Threshold model. Its articulated mood is calculated from Equation. (5), wherein $\beta_j \in$ [0, 1] is a variable defining the relative relevance of intrinsic sentiment to neighbouring behaviour.

$$FV_j[B] = (1-\beta_j).FV_j[IB]+\beta_j. \sum_{v_i \in BN_j} FV_i[B].RE_{ij}[P].RE_{ij} \tag{5}$$

The significance of clustering structure on the impact of a collection in weighted networks is examined by Mahsa et al. [24]. The researchers use a game theory technique to replicate the dissemination phase in the network for this goal. The relationship between the effect of distinct sets and positive inter-community links and intra-community negative links are evaluated. They came to the conclusion that there may be a link among a seed set effectiveness and the population.The more intra-community negative relationships there are, the more complex the system becomes therefore increasing the influencing capacity of the seeds .Yanhua et al. [27] proposes a weighted voting system to mimic dissemination in weighted networking graphs. Each vertex in this system has a favourable or unfavourable opinion. In the beginning, when time = 0, all nodes are good. Every vertex v_i picks a few of its oriented neighbours, such as v_j, with such a likelihood due to the value of the link linking these vertices at every epoch t > 0. If the link's value is greater than 0, v_i

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

adjusts its viewpoint to match $v_j's$, that is, $FV_i[O] = FV_j[O]$, however it adjusts its viewpoint to match the contrary of $v_j's$, that is, $FV_i[O] = FV_j[O]$. Seed nodes are usually engaged and have either a favourable or unfavourable attitude; all the other vertices are dormant and impartial. Every vertex v_i that is triggered in t =1 has a potential to engage every dormant neighbor v_j having initiation probability RE_{ij}. If user v_j is enabled, its view is $FV_j[O] = FV_i[O].RE_{ij}[P]$. Table 1 summarises the attributes of the described attitude-dependent approaches.

TABLE 1: Features of attitude-dependent techniques, including characteristics and technique for ranking influential individuals: Centrality Measure(I_1), Spreading Simulation (I_2), Reverse-Influence Sampling(I_3),and Maximum-Influence Aborescence(I_4)

Type	Paper No.	Characteristics					Influence detection			
		C2	C8	C9	C10		I_1	I_2	I_3	I_4
Trust-independent	[19]	✓		✓	✓		✓			
	[40]	✓		✓				✓		
	[36]	✓		✓				✓		
	[9]		✓					✓		✓
	[32]		✓				✓	✓		
Trust-dependent	[27]	✓	✓				✓			
	[7]	✓		✓				✓		
	[34]	✓		✓	✓			✓		
	[31]	✓		✓			✓			
	[22]		✓				✓			
	[33]	✓	✓					✓		
	[4]	✓	✓							
	[15]	✓	✓				✓	✓		
	[16]	✓	✓					✓		
	[29]	✓	✓				✓	✓		
	[14]	✓	✓				✓			
	[8]	✓	✓					✓		
	[13]	✓	✓	✓				✓		

Description of Characteristics:	
C2	$RE_{ij} \in [0,1]$ describes the dissemination probability of link e_{ij}.
C8	$FV_i[O] \in \{1,0,+1\}$ or {agree,disagree} represents the attitude of the node v_i towards the information.
C9	$FV_i[B] \in [0,1] or [1,+1]$ represents the attitude of node v_i towards the data; attitude alters during the dissemination process
C10	$FV_i[IB] \in [0,1]$ represents the inherent attitude of node v_i towards the data; attitude that doesn't alter

V. MONEY-DEPENDENT TECHNIQUES

The financial features of the spreading process are kept in mind in such techniques. Participants are compensated in order to motivate participants to participate in the seed-spreading procedure. Furthermore, certain participants may be activated during the ranking of users. There are financial rewards. Money-dependent techniques are categorised into the following groups:

*1) **Budgeted-Ranking User Technique**:* In such techniques, picking every individual vi as a highest-ranking individual has a price, i.e., $FV_i[p] \in [0,1]$ refers to the price of choosing v_i as a highest ranking individual. Additionally, there is an overall budget-Quota, B, which should not be exceeded by the summation of seed prices. It means, rather than supplying a certain amount of highest-ranking individuals as in earlier techniques, query in such techniques must be guided by a budget-Quota to find the highest-ranking individuals. Usually Budgeted- Ranking users research assumes that components with more companions are extra expensive to convince and participate in distribution technique. This indicates the price of utilising components as the highest-ranking individual is proportional to the links it has with other components.

Nguyen et al. [5] were the first to develop a budgeted-Ranking user system. They used the following equation to represent the ranking users issue, where P(I) is the total price of all highest ranking individuals, $P(I) = \sum_{v_i \in I} FV_i[P]$.

$$I^* = argmax_{I \subset V} f(I) \quad subjecttoP(I) < B \tag{6}$$

They presented a greedy technique for finding the highest-ranking individuals in multiple rounds based on distributing analysis underneath the IC diffusion model [23]. The component with the greatest effective impact for price gets included into the highest-ranking individuals groups with every iteration. They further developed an updated greedy technique to tackle this issue, wherein researchers employed Maximal Impact Pathways (MIPs) to build a directed-acyclic-graph and applied a belief propagation algorithm [47] to estimate the distributing capability of components.

In Suman et al. [18] the graph is separated into regions and depending on the number of nodes in each region, the budget is distributed across the various regions. Small regions are ranked first, followed by bigger regions. Components are ordered in non-increasing order on the basis of the links it has with other components. In rounds, components from every region are included in highest-ranking individual groups, with the leading component of the sequence (sorted) considered a contender in every repetition. If somehow the price of including a contender component is less than the budget quota left, then it will get included in the highest-ranking individual groups as the latest individual; else, the following component of the list is selected as a contender. Repetitions proceeds till the region's money is depleted or the group cannot expand. If funds are left, they are moved to a wider region with more intra-community linkages to the present one. The procedure is iterated for remainder regions to discover the highest-ranking individual.

*2) **Gain-Maximisation Methods** :* Besides the financial expenses of persuading an individual to participate as the highest-ranking individual of the distribution procedure, every individual's excitation might provide an advantage. A financial gain $FV_i[B]$ if it is turned on while the virus is propagating. Whenever this situation occurs, the aim is to maximise total gain depending upon price gain of distributing procedure in Nguyen et al. [35] Every individual v_i contains $FV_i[C]$ and $FV_i[B]$ reflecting the price if it gets included in highest-ranking individuals group, and the profit of activating it throughout the distribution phase, respectively. The ranking users issue is described as finding the highest-ranking users with a price that is not bigger than B, thus optimising the overall profit from all active components at the same time. The technique of Reverse Influence Sampling (RIS) could then be applied to locate the impactful individuals. Components with a bigger effect on Components that pay out high profits are expected to be selected in the highest-ranking individuals. Distribution is induced in Huiyuan et al. [39], where the goal is to increase the active individuals for various items.

The objective is to get the most out of all of the active

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

TABLE 2: Features of money-dependent techniques, including characteristics and technique for ranking influential individuals: Centrality Measure (I_1), Spreading Simulation (I_2), Reverse Influence Sampling (I_3), and Maximum Influence Aborescence (I_4)

Type	Paper No.	Characteristics						Influence detection			
		C2	C11	C12	C13	C14	C15	I_1	I_2	I_3	I_4
Budgeted Ranking	[5]	✓	✓						✓		✓
	[11]		✓					✓			
	[18]	✓	✓					✓			
	[46]	✓	✓					✓			
	[30]	✓	✓					✓			
	[3]	✓	✓						✓		
	[21]	✓	✓					✓			
	[6]	✓	✓					✓			
Gain-maximization	[35]	✓	✓	✓						✓	
	[39]	✓	✓	✓			✓	✓	✓		
	[17]	✓	✓		✓	✓			✓		
	[20]	✓	✓	✓					✓		
	[38]	✓	✓	✓						✓	
Description of Characteristics:											
C2	$RE_{ij} \in [0,1]$ describes the dissemination probability of link e_{ij}.										
C11	$FV_i[C] \in [0,1]$ represents the price of convincing node v_i to participate in the dissemination process as a highest ranking node										
C12	$FV_i[B] \in [0,1]$ represents the advantage from node v_i, if the node is engaged throughout dissemination process										
C13	$FV_i[P] \in [0,1]$ represents the cost recommended to node v_i in the dissemination process										
C14	$FV_i[V] \in [0,1]$ represents node $v_i's$ estimate about the item broadcast by search-query										
C15	QF[D] a group of items broadcast by the search-query.										

individuals and all of the items. Furthermore, to mimic the distribution procedure wherein component v_j shifts to exciting phase for items p if it satisfies some conditions, an adaptation of the LT diffusion model [37] is given. The technique is employed repeatedly, with the component with the largest proportion of impact to the price being included to the highest-ranking individuals as a newer individual in every repetition. Because simulation-based approaches are inefficient, the researchers present several heuristics algorithms that have good time efficiency. Researchers also describe the issue as a multi-choice knapsack problem to divide the budget quota between the items as efficiently as possible and then analyse it using binary integer computing. The aim in Wei Lu et al. [17, 20] is to maximise gain, i.e., gain -price, with no budget-quota constraints. With each component v_i, a cost $FV_i[P]$ for an item is examined in Wei Lu et al. [17], and the LT diffusion model is then expanded as below. Every component could be in one of two states: active (adopted) or dormant (adopted). The impacting procedure is like in the traditional LT diffusion model, except if the item's value, $FV_i[V]$, is larger than its pricing, $FV_i[P]$, the affected component vi shifts to the chosen form. Then, using the suggested diffusion model, a greedy method is developed to locate the individuals that maximise gain. On the basis of the price of convincing the individuals of the original highest-ranking individuals and the gain of the active components throughout the distributing phase, the degree discount method is enhanced in Jing Tang et al.[20] to discover impactful highest-ranking individuals for maximum gain.

Table 2 summarises the attributes of the described money-dependent approaches.

VI. Challenges and Future Directions

We offered a unique categorisation of approaches presented to tackle the Ranking Users problem in social networks in this paper, emphasising solutions that include characteristics of individuals' behaviour. There is a lot of study going on concerning the Ranking Users problem, and there is a massive curiosity because there are so many obstacles to overcome. As Impartial behaviour algorithms employ a mechanistic technique to connect people over a network, in our opinion, they are often not accurate sufficient to find impactful components in practical uses. Although it's been suggested that the internet can reveal an individual's characteristics, physical-world individual behaviour is certainly considerably greater complicated than what a network can record. Another issue that has to be addressed in further studies is indeed the scale and variety of the network. The amount of individuals and the amount of information transmitted among them is growing with each day. Lastly, a diffusion model is used to represent the stream of data in practically all approaches. Such methods might not be the most accurate representations of how knowledge spreads in practical systems. The precision of such approaches could be impacted if the diffusion model is changed. To produce better and more accurate diffusion models, more study is needed to examine the influence of individuals' behaviour in the distributing processes.

References

[1] S.Kumar, A. Mallik, and B. S. Panda. "Link prediction in complex networks using node centrality and light gradient boosting machine." World Wide Web (2022): 1-27.

[2] S.Kumar, A. Gupta, and I. Khatri. "CSR: A community based spreaders ranking algorithm for influence maximization in social networks." World Wide Web (2022): 1-20.

[3] Evren G¨uney. "On the optimal solution of budgeted influence maximization problem in social networks". In: Operational Research 19.3 (2019), pp. 817–831.

[4] Dong Li et al. "Positive influence maximization in signed social networks based on simulated annealing". In: Neurocomputing 260 (2017), pp. 69– 78.

[5] Huy Nguyen and Rong Zheng. "On budgeted influence maximization in social networks". In: IEEE Journal on Selected Areas in Communications 31.6 (2013), pp. 1084–1094.

[6] RC de Souza et al. "Efficient Network Seeding under Variable Node Cost and Limited Budget for Social Networks". In: Information Sciences 514 (2020), pp. 369–384.

[7] Huanhuan Wang et al. "Maximizing positive influence in signed social networks". In: International Conference on Cloud Computing and Security. Springer. 2015, pp. 356–367.

[8] Sabbir Ahmed and CI Ezeife. "Discovering influential nodes from trust network". In: Proceedings of the 28th annual ACM symposium on applied computing. ACM. 2013, pp. 121–128.

[9] Wei Chen et al. "Influence maximization in social networks when negative opinions may emerge and propagate". In: Proceedings of the 2011 SIAM international conference on data mining. SIAM. 2011, pp. 379–390

[10] Linton C Freeman. "A set of measures of centrality based on betweenness". In: Sociometry (1977), pp. 35–41.

[11] Shuo Han et al. "Balanced seed selection for budgeted influence maximization in social networks". In: Pacific-Asia Conference on Knowledge Discovery and Data Mining. Springer. 2014, pp. 65–77.

[12] Linyuan L¨u et al. "The H-index of a network node and its relation to degree and coreness". In: Nature communications 7 (2016), p. 10168.

[13] Feng Wang, Guojun Wang, and Dongqing Xie. "Maximizing the spread of positive influence under LT-MLA model". In:

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

Asia-Pacific Services Computing Conference. Springer. 2016, pp. 450–463.

[14] Jing Selena He, Harneet Kaur, and Manasvi Talluri. “Positive opinion influential node set selection for social networks: Considering both positive and negative relationships”. In: Wireless Communications, Networking and Applications. Springer, 2016, pp. 935–948.

[15] Ajitesh Srivastava, Charalampos Chelmis, and Viktor K Prasanna. “Social influence computation and maximization in signed networks with competing cascades”. In: 2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM). IEEE. 2015, pp. 41–48.

[16] Maryam Hosseini-Pozveh et al. “Maximizing the spread of positive influence in signed social networks”. In: Intelligent Data Analysis 20.1 (2016), pp. 199–218.

[17] Wei Lu and Laks VS Lakshmanan. “Profit maximization over social networks”. In: 2012 IEEE 12th International Conference on Data Mining. IEEE. 2012, pp. 479–488.

[18] Suman Banerjee, Mamata Jenamani, and Dilip Kumar Pratihar. “ComBIM: A community-based solution approach for the Budgeted Influence Maximization Problem”. In: Expert Systems with Applications 125 (2019), pp. 1–13.

[19] Aristides Gionis, Evimaria Terzi, and Panayiotis Tsaparas. “Opinion maximization in social networks”. In: Proceedings of the 2013 SIAM International Conference on Data Mining. SIAM. 2013, pp. 387–395.

[20] Jing Tang, Xueyan Tang, and Junsong Yuan. “Profit maximization for viral marketing in online social networks: Algorithms and analysis”. In: IEEE Transactions on Knowledge and Data Engineering 30.6 (2017), pp. 1095–1108.

[21] Yufeng Wang et al. “PPRank: economically selecting initial users for influence maximization in social networks”. In: IEEE Systems Journal 11.4 (2015), pp. 2279–2290.

[22] Shubo Chen and Kejing He. “Influence maximization on signed social networks with integrated pagerank”. In: 2015 IEEE International Conference on Smart. IEEE. 2015, pp. 289–292.

[23] David Kempe, Jon Kleinberg, and Eva Tardos. “Maximizing the spread ´ of influence through a social network”. In: Proceedings of the 9th ACM SIGKDD international conference on Knowledge discovery and data mining. ACM. 2003, pp. 137–146.

[24] Mahsa Shafaei and Mahdi Jalili. “Community structure and information cascade in signed networks”. In: New Generation Computing 32.3-4 (2014), pp. 257–269.

[25] Kumar, Sanjay, Dipti Lohia, Darsh Pratap, Ashutosh Krishna, and B. S. Panda. ”MDER: modified degree with exclusion ratio algorithm for influence maximisation in social networks.” Computing (2021): 1-24.

[26] Maksim Kitsak et al. “Identification of influential spreaders in complex networks”. In: Nature physics 6.11 (2010), p. 888.

[27] Yanhua Li et al. “Influence diffusion dynamics and influence maximization in social networks with friend and foe relationships”. In: Proceedings of the 6th ACM international conference on Web search and data mining. ACM. 2013, pp. 657–666.

[28] Mahdi Jalili and Matja˘z Perc. “Information cascades in complex networks”. In: Journal of Complex Networks 5.5 (2017), pp. 665–693.

[29] Rezvan Mohamadi-Baghmolaei, Niloofar Mozafari, and Ali Hamzeh. “Trust based latency aware influence maximization in social networks”. In: Engineering Applications of Artificial Intelligence 41 (2015), pp. 195–206.

[30] Nan Du et al. “Budgeted influence maximization for multiple products”. In: arXiv preprint arXiv:1312.2164 (2013).

[31] S. Kumar, M. Saini, M. Goel, and B. S. Panda. ”Modeling information diffusion in online social networks using a modified forest-fire model.” Journal of intelligent information systems 56, no. 2 (2021): 355-377.

[32] A. Bhowmik, S.Kumar, N. Bhat. ”Evolution of automatic visual description techniques-a methodological survey.” Multimedia Tools and Applications 80, no. 18 (2021): 28015-28059.

[33] S. Kumar, M. Saini, M. Goel, and N. Aggarwal. ”Modeling information diffusion in online social networks using SEI epidemic model.” Procedia Computer Science 171 (2020): 672-678.

[34] S. Kumar, L. Singhla, K. Jindal, K. Grover, and B. S. Panda. ”IM-ELPR: Influence maximization in social networks using label propagation based community structure.” Applied Intelligence 51, no. 11 (2021): 7647-7665.

[35] Hung T Nguyen, Thang N Dinh, and My T Thai. “Cost-aware targeted viral marketing in billion-scale networks”. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE. 2016, pp. 1–9.

[36] S. Kumar, B. S. Panda, and D. Aggarwal. ”Community detection in complex networks using network embedding and gravitational search algorithm.” Journal of Intelligent Information Systems 57, no. 1 (2021): 51-72.

[37] Allan Borodin, Yuval Filmus, and Joel Oren. “Threshold models for competitive influence in social networks”. In: International workshop on internet and network economics. Springer. 2010, pp. 539–550.

[38] Jianming Zhu et al. “Profit Maximization Under Group Influence Model in Social Networks”. In: International Conference on Computational Data and Social Networks. Springer. 2019, pp. 108–119.

[39] Huiyuan Zhang et al. “Profit maximization for multiple products in online social networks”. In: IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. IEEE. 2016, pp. 1–9.

[40] Huiyuan Zhang, Thang N Dinh, and My T Thai. “Maximizing the spread of positive influence in online social networks”. In: 2013 IEEE 33rd International Conference on Distributed Computing Systems. IEEE. 2013, pp. 317–326.

[41] An Zeng and Cheng-Jun Zhang. “Ranking spreaders by decomposing complex networks”. In: Physics Letters A 377.14 (2013), pp. 1031–1035.

[42] Ahmad Zareie, Amir Sheikhahmadi, and Mahdi Jalili. “Influential node ranking in social networks based on neighborhood diversity”. In: Future Generation Computer Systems 94 (2019), pp. 120–129.

[43] Ahmad Zareie, Amir Sheikhahmadi, and Adel Fatemi. “Influential nodes ranking in complex networks: An entropy-based approach”. In: Chaos, Solitons Fractals 104 (2017), pp. 485–494.

[44] S. Kumar, and B. S. Panda. ”Identifying influential nodes in Social Networks: Neighborhood Coreness based voting approach.” Physica A: Statistical Mechanics and its Applications 553 (2020): 124215.

[45] S. Kumar, and A. Panda. ”Identifying influential nodes in weighted complex networks using an improved WVoteRank approach.” Applied Intelligence (2021): 1-15.

[46] Qilian Yu et al. “Fast budgeted influence maximization over multi-action event logs”. In: IEEE Access 6 (2018), pp. 14367–14378.

[47] Jonathan S Yedidia, William T Freeman, and Yair Weiss. “Understanding belief propagation and its generalizations”. In: Exploring artificial intelligence in the new millennium 8 (2003)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:11:45 UTC from IEEE Xplore. Restrictions apply.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI: 10.17775/CSEEJPES.2021.07310, CSEE Journal of Power and Energy Systems

Reweighted L_0 norm variable step size continuous mixed p-norm control scheme for mitigating power quality problems of grid coupled solar PV system

Pallavi Verma, *Student Member, IEEE*, Avdhesh Kumar, *Student Member, IEEE* , Rachana Garg, Senior *Member, IEEE* and Priya Mahajan, Senior *Member, IEEE*

Abstract— **In this paper, authors analyzed the performance of two stage-three phase grid coupled solar photovoltaic generating system (SPVGS) by using novel reweighted L_0 norm variable step size continuous mixed p-norm (RL_0-VSSCMPN) of voltage source inverter (VSI) control scheme. The efficacy of the system is determined by considering unbalanced grid voltage, dc offset, voltage sag and swell, unbalanced load and variation in solar insolation. RL_0-VSSCMPN is used for inverter control and it extracts fundamental component of load current for generating reference grid current with faster convergence rate and lesser steady state oscillations. With the proposed control, harmonics in grid current follows the IEEE-519 norm and the performance is also satisfactory under varying environmental/load conditions. The power generated from SPVGS is transferred optimally using dc-dc boost converter utilizing incremental conductance (INC) maximum power point technique. The proposed system is simulated using MATLAB/Simulink 2018a and test results are verified experimentally using dSPACE1202 in the laboratory to ensure the validity of novel proposed robust RL_0-VSSCMPN.**

***Index Terms*-- INC maximum power point tracker, power quality, reweighted L_0 VSSCMPN algorithm, solar PV generating system, total harmonic distortion, voltage source inverter.**

I. INTRODUCTION

Solar photovoltaic generating system is one of the most important green source of energy. Stand-alone as well as grid coupled system are keenly accommodating this transformation of energy which is a big boon and positive step for SPVGS. But the solar energy is available only when the sun shines and the maximum power point unceasingly fluctuates due to weather condition. So, to extract maximum power from solar photovoltaic system, maximum power point tracking (MPPT) is used [1], [2]. Many MPPT control algorithms are reported in literature like perturb and observation, incremental conductance, adaptive neuro fuzzy system, genetic algorithm etc. [3]–[6].

Solid state device-based loads have made our life easier but they also draw harmonics from the utility grid and raises power quality issues like poor power factor, unbalancing etc. Grid coupled SPVGS also necessitate these power electronics components. Voltage source inverters are used to mitigate the power quality issues. So, an efficient VSI control technique becomes imperative as it reduces the reactive power demand from utility grid and reduces harmonics

Some of the VSI control schemes reported in literature are synchronous rotating frame theory (SRFT), unit template, *dq* control theory, instantaneous reference frame, second order generalized integrator (SOGI), adaptive based control and are presented in both power factor correction (PFC) as well as zero voltage regulation mode [7]–[12]. Phase lock loop (PLL) is used to convert three phase quantities into two phase in SRFT control algorithm. But it requires proper tuning before implementation. There are many modification of PLL reported in literature [13]–[16] which shows zero steady state oscillations but are incapable of rejecting dc offset component. A. Mortezaei [17] presented comparative analysis of *pq*, *dq* and conservative power theory (CPT) for shunt active compensators. The *dq* provides pure sinusoidal grid current under distorted current condition. Also, the *dq* has faster response than other presented techniques but it fails to reject the ripples in current for unbalanced load. The SOGI control can estimate the fundamental component of load but is unable to discard the direct current offset in load current. Furthermore, the systems performance under distorted grid voltage has not been considered. Various artificial intelligence-based PV inverter control algorithm are also reported. S. Nanda et al. in [18] gives new control adapting filter based on modified Gauss–Newton (MGNA) for calculating fundamental and harmonic phasors of non - linear power system waveforms as well as frequency change. N. Gupta et al. [19] implemented fuzzy controller for PV inverter control of grid connected SPV system. The membership function of the fuzzy are asymmetrically distributed which helps in early settling of dc-link voltage as compared to other conventional methods. The total harmonics distortion (THD) is within IEEE norms under steady state and dynamic situations. Authors in [20] have developed an intelligent controller named as Takagi-Sugeno-Kang type Probabilistic FNN having an Asymmetric Membership Function (TSKPFNN-AMF) for three phase grid integrated PV system. Authors have discussed the network structure, online learning algorithms, and convergence of the developed controller. Further the performance of TSKPFNN-AMF has been compared with PI, PID, FNN and WFNN techniques. But these techniques require proper tuning before its implementation.

Many adaptive control schemes such as least mean square (LMS), least mean fourth (LMF), zero attracting LMS,

improved, proportionate normalized LMS, reweighted L_1 (RL_1) norm penalized LMSF [12], [21], [22] are used by various researchers for power quality refinement. Better performance under steady state condition can be attained using LMS while LMF has faster convergence rate. B. Sun in [23] proposes a filtering method entitled self-adaptive noise cancellation (SANC) to minimize the effect of vibrations, as well as a novel variable step-size least mean square (NVS-LMS) algorithm for transporter belt-type checkweigher. Authors in [24] mitigates the power quality issues of shunt active power filter. The Wiener filter-based control scheme is used to extract the fundamental load component and the performance is compared with LMS. The performance of LMS was found to be unsatisfactory under dynamic condition. In [25] author's presented Kernel Hebbian least mean square (KHLMS) for controlling the DSTATCOM. It is the improved version of adaptive LMS and on comparison it is observed that the KHLMS algorithm has better DC link response, reduced harmonic in source current, balanced 3-phase voltages etc. Authors in [21] presented variable step size LMS (VSSLMS) for detecting current harmonics. The VSSLMS has provided adjustment control with faster convergence rate. R. K. Agarwal et al. [26] reported leaky LMF for SPV-DTSATCOM. Results have been validated on both simulation and hardware. The performance is satisfactory for both steady state and dynamic conditions but the dc-offset rejection capability has not been considered. The introduction of reweighted zero attracting component to LMSF makes convergence speed faster having low error. In [27] authors have implemented functional link artificial neural network (FLANN) for shunt compensation. The developed algorithm performs satisfactorily under steady and transient response as well as it has fast convergence. In [28] reweighted zero attracting quaternion LMS (ZAQLMS) for distributed generation system with induction motor load was presented. The proposed control scheme rejects the dc offset from load current and also mitigates other PQ issues. The performance of conventional adaptive algorithm suffers under grid distorted voltage, dc offset and imbalanced voltage. Robust mixed norm (RMN) was discussed in [29]. Detailed analysis of p-norm and mixed p-norm were reported in [11], [30], [31]. A. Ranjan et al. [32] investigated reweighted L_1 norm least mean fourth (RL_1-LMSF) with improved multivariable filter based frequency locked loop for grid connected PV system. The control scheme is tested for abnormal grid condition as well as load condition. Test results confirms that the proposed algorithm works satisfactorily under all the scenarios. Authors [33] presented improved performance of grid coupled PV system by continuous mixed p-norm (CMPN). The comparative analysis of CMPN with conventional LMS control algorithm was presented.

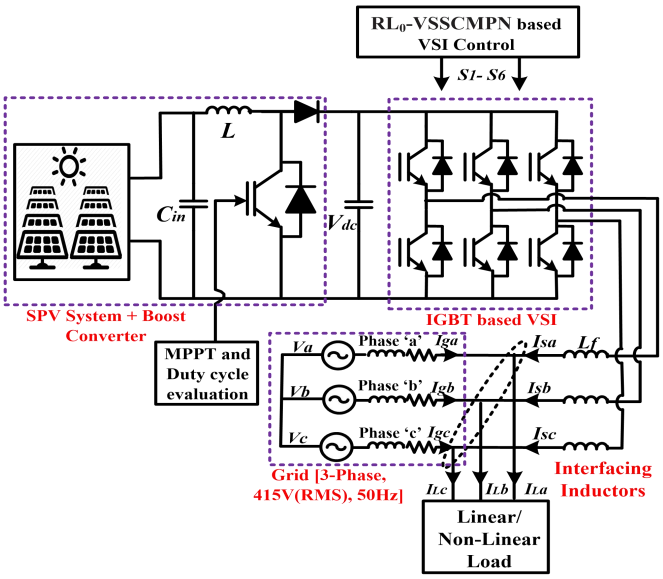


Fig. 1. Schematic of the grid coupled solar PV generating system

Recent research works for VSI control emphasizes upon LMS based algorithm due to its ease of application, but impulsive noise generated by LMS algorithm deteriorates the overall system's performance. So, few advanced p-norm adaptive filter algorithms have been introduced and discussed above. The fixed step-size in continuous mixed p-norm makes the algorithm parameter dependent. Here, reweighting component is added which gives fast convergence speed and makes system adaptable to complicated circuits The variable step size - CMPN algorithm is used to act on this constraint. Fixed step size in CMPN is also made variable which changes in accordance to change in input/output parameters. Authors have implemented optimal novel controller reweighted L_0 norm combined with variable step size CMPN (RL_0-VSSCMPN). In this paper, grid coupled SPVGS under grid abnormalities and input/output variation is dealt with proposed RL_0-VSSCMPN control scheme. The key aspects of the presented research contribute towards:

- Development of an efficient control scheme for grid coupled SPVGS. Grid current remains sinusoidal under all the scenarios and balanced following IEEE-519 standard.
- The reweighted L_0-VSSCMPN algorithm shares the generated power from solar with load and grid.
- The power quality issues like voltage sag, swell, load unbalancing, grid unbalancing, dc-offset etc. are addressed.

Improved RL_0-VSSCMPN control algorithm is less complex, has less steady-state error, minimum oscillations in dynamic condition, faster tracking and convergence speed, and has capability of dc-offset rejection. The efficacy of the proposed control scheme under dynamic situation viz. dc offset, voltage sag, swell, change in solar irradiation, unbalanced non-linear load has been tested on both simulation and hardware prototype developed in the laboratory. The obtained results are within the IEEE-519 standard.

To mitigate the PQ issues raised because of grid integration of

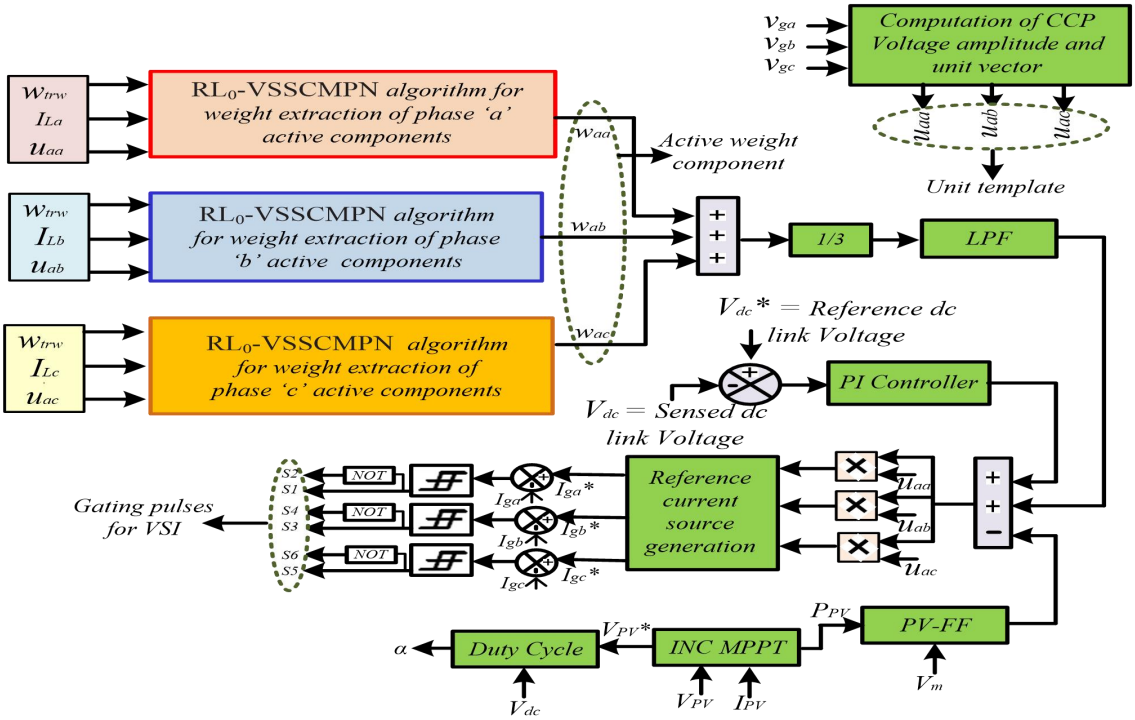


Fig. 2. Reweighted L_0-VSSCMPN control scheme

RES using reweighted L_0 VSSCMPN algorithm for PV inverter grid coupled SPVGS is developed in MATLAB/Simulink as presented in Section II. In Section III, the detailed explanation of the proposed control schemes used for MPPT and VSI are given. Section IV presents the results obtained from the MATLAB/Simulink under steady as well as dynamic condition. In Section V, results obtained from hardware prototype has been presented. Section VI completes the findings of the proposed work.

II. Layout of The Proposed System

The layout of the proposed system is shown in Fig. 1. It has double stage solar photovoltaic (PV) generating system, 3ϕ-415V (RMS), 50Hz utility grid and voltage source inverter (VSI) to connect SPVGS, load and grid. The SPVGS having rating of 11.78kW generates DC output, which is connected to boost converter to produce regulated DC output. DC-DC boost converter acts as an interface between solar PV array and dc-link capacitor (V_{dc}) of VSI. Nonlinear load is connected at the common coupling point (CCP) using interfacing inductors (L_f). Voltage source inverter are used to convert regulated DC output into AC voltage and also serves the functions of harmonics elimination, reactive power compensation etc. Two control schemes are used in the present work: firstly- INC MPPT control algorithm is used to extract the maximum power from PV array and second: proposed reweighted L_0 continuous mixed p-norm adaptive filtering control scheme is used for gating the six IGBT switches of VSI.

III. Control Scheme Used in The Proposed System

The VSI control scheme used in grid coupled solar PV generating system is shown in Fig. 2. The INC based MPPT is used to track peak power point of solar PV array. Main objective of VSI control scheme is to eliminate harmonics and balanced grid current at unity power factor. The reference grid current is computed by PV feed-forward, active loss term and load currents fundamental component. Hysteresis current controller is used for generating gating pulses for IGBT switches of VSI.

The implemented control scheme is presented in two subsections: ***A.*** maximum power extraction and ***B.*** VSI control scheme and these are explained below:

A. Maximum Power Extraction

In the present paper, INC MPPT is used to track the maximum power available from solar PV array and made it to operates at its peak point using boost converter.

The duty cycle at i^{th} instant is evaluated using voltage across dc-link and reference PV voltage which can be expressed as given below:

$$D(i) = 1 - \frac{V^*_{PV}(i)}{V_{dc}(i)} \quad (1)$$

where, V^*_{PV} is reference PV voltage and V_{dc} is voltage across dc-link. Step size in the implemented algorithm is 0.001.

B. VSI control scheme

In the present work, the proposed control scheme RL_0-VSSCMPN was implemented for alleviating the power quality problems. The active weight components are calculated using proposed control scheme. The average weight of fundamental load currents is calculated using equation (11) and total reference grid current is obtained by equation (15).

The VSI control scheme must hold two features to achieve sinusoidal and balanced grid current. 1) Extraction of load currents fundamental component 2) calculation of unit template. Moreover, it compensates the reactive need of the load and also reduces the harmonics of the grid current. For reliable and effective operation of the proposed control voltages of dc-link (V_{dc}), PV array (V_{PV}) and common coupling point ($V_{gab}, V_{gbc}, V_{gca}$) and currents of the grid (I_{ga}, I_{gb}, I_{gc}), the load (I_{La}, I_{Lb}, I_{Lc}) and the PV array (I_{PV}) are sensed.

1) Grid current reference magnitude estimation

The grid currents reference is approximated by weights of the fundamental load current components, feed-forward PV term and active loss component.

The RL_0-VSSCMPN control algorithm has been used to estimate the fundamental load current component. It evaluates the weight (I_{wa}, I_{wb}, I_{wc}) for each phase of the load current. To obtain balanced grid current, the average of these weights is used to evaluate reference grid currents.

1) Estimation of reference grid current

The equations given below are involved in RL_0-VSSCMPN control algorithm.

The cost function $C_{RL_0}(i)$ with reweighted L_0 norm function of the algorithm is given as [34]:

$$C_{RL_0}(i) = \int_1^2 \lambda_i(p)\, E\{|e(i)|\}^p \mathrm{d}p + \underbrace{\gamma \|f(i)\, w(i)\|_0}_{reweighted L_0 norm} \quad (2)$$

where, $\lambda_i(p)$ is a probability density-like weighting function with the constraint $\int_1^2 \lambda_i(p)\mathrm{d}p = 1$, $\|.\|_0$ is L_0 norm, γ is regularization parameter makes convergence faster and smoothens transitions.

The error function of phase 'a' load is given by:

$$e(i) = I_{la}(i) - I_{wa}(i)u_{aa}(i) \quad (3)$$

$$f(i) = \frac{1}{\epsilon + |I_{wa}(i-1)|} \quad (4)$$

where u_{aa} is unit template, $\epsilon > 0$ and $w(i)$ weight value.

Weight extraction by RL_0-VSSCMPN

The weight (I_{wa}) of fundamental load current component is given by:

$$I_{wa}(i) = I_{wa}(i-1) + \mu \frac{\partial C_{RL_0}(i-1)}{\partial I_{wa}(i-1)} \quad (5)$$

$$I_{wa}(i) = I_{wa}(i-1) + \mu \int_1^2 \lambda_{i-1}(p) \frac{\partial E\{|e(i-1)|\}^p}{I_{wa}(i-1)} \mathrm{d}p - \mu\, \gamma\, f_\beta\{f(i-1)\, w(i-1)\} \quad (6)$$

$$I_{wa}(i) = I_{wa}(i-1) + \mu\, \mathfrak{Z}_{i-1} sgn\big(e(i-1)\big) x(i-1) - \mu\, \gamma\, f_\beta\{f(i-1)\, w(i-1)\} \quad (7)$$

where μ is the step size, *sgn* is signum function, $\mathfrak{Z}_{i-1} = \int_1^2 p\lambda_i(p)\, |e(i)|^{p-1}\mathrm{d}p$ is variable step size that depends on e(i).

$$\mathfrak{Z}_{i-1} = \frac{(2|e(i)|-1)ln(|e(i)|) - |e(i)|+1}{ln^2(|e(i)|)} \quad (8)$$

Similarly, the fundamental component of weight for other two phases are given below [30], [32], [33]:

$$I_{wb}(i) = I_{wb}(i-1) + \mu\, \mathfrak{Z}_{i-1} sgn\big(e(i-1)\big) x(i-1) - \mu\, \gamma\, f_\beta\{f(i-1)\, w(i-1)\} \quad (9)$$

$$I_{wc}(i) = I_{wc}(i-1) + \mu\, \mathfrak{Z}_{i-1} sgn\big(e(i-1)\big) x(i-1) - \mu\, \gamma\, f_\beta\{f(i-1)\, w(i-1)\} \quad (10)$$

The average weight of fundamental load currents is given by:

$$I_{wmean}(i) = \frac{I_{wa}(i) + I_{wb}(i) + I_{wc}(i)}{3} \quad (11)$$

This (I_{wmean}) is given to low pass filter which generates fundamental load current represented as (I_{fwmean}).

The active loss component ((I_{loss}) which regulates the voltage across DC bus is estimated by using v_{de} and a PI controller gains as:

$$I_{loss}(i) = w_{dc}(i-1) + K_{pd}\{v_{de}(i) - v_{de}(i-1)\} + K_{id} v_{de}(i) \quad (12)$$

where, $K_{pd} and\ K_{id}$ are PI controller proportional and integral gains respectively, v_{de} is error generated by comparing sensed dc-link (V_{dc}^*) and reference dc link voltage (V_{dc}) voltages and the generated error is given to PI controller.

$$v_{de}(i) = V_{dc}^*(\mathrm{i}) - V_{dc}(\mathrm{i}) \quad (13)$$

A feed-forward PV (I_{PV-ff}) term improves the performance at voltage and current disturbances which is given by:

$$I_{PV-ff} = \frac{2\, P_{PV}}{3 V_m} \quad (14)$$

where V_m is the maximum voltage magnitude at the coupling nodes and P_{PV} is the instantaneous PV power.

Hence, total reference grid current (I_{total}) is given by:

$$I_{total} = I_{wmean} + I_{loss} - I_{PV-ff} \quad (15)$$

2) Estimation of unit templates: Sinusoidal grid currents is achieved by accurate estimation of unit templates. The point of common coupling (PCC) line voltages (V_{gab}, V_{gbc}) is used to compute the PCC phase voltage (V_a, V_b, V_c) and its peak amplitude are given below.

$$\begin{bmatrix} V_a \\ V_b \\ V_c \end{bmatrix} = \frac{1}{3}\begin{bmatrix} -2 & 1 \\ -1 & 1 \\ -1 & -2 \end{bmatrix}\begin{bmatrix} V_{gab} \\ V_{gbc} \end{bmatrix} \quad (16)$$

$$V_t = \sqrt{\frac{2}{3}\left(V_a^2 + V_b^2 + V_c^2\right)} \quad (17)$$

The in-phase unit templates (u_{aa}, u_{ab}, u_{ac}) are given by:

$$u_{aa} = \frac{V_a}{V_t}, u_{ab} = \frac{V_b}{V_t}, u_{ac} = \frac{V_c}{V_t} \quad (18)$$

3) Pulse generation for gating VSI switches:

The reference grid currents ($i_{ga}^*, i_{gb}^*, i_{gc}^*$) are computed by multiplication of reference grid current with unit templates which are given by:

$$i_{ga}^*(i) = I_{total}(i) * u_{aa}(i), i_{gb}^*(i) = I_{total}(i) * u_{ab}(i), i_{gc}^*(i) = I_{total}(i) * u_{ac}(i) \quad (19)$$

The current error (i_{ea}, i_{eb}, i_{ec}) is given by:

$$i_{ea}(i) = i_{ga}^*(i) - i_{ga}(i), i_{eb}(i) = i_{gb}^*(i) - i_{gb}(i) \quad i_{ec}(i) = i_{gc}^*(i) - i_{gc}(i) \quad (20)$$

The generated error signals are given to hysteresis current controller for providing gating pulses to VSI switches.

The control signal flow diagram for the proposed system as shown in Fig.3, in which various parameters viz. DC link voltage, load current, grid current and PCC voltage are sensed at PCC. Further, the active weight component, feed forward PV term and total reference grid current are calculated using the sensed parameters. In last, the actual and reference grid current are compared to generate gating signal for VSI.

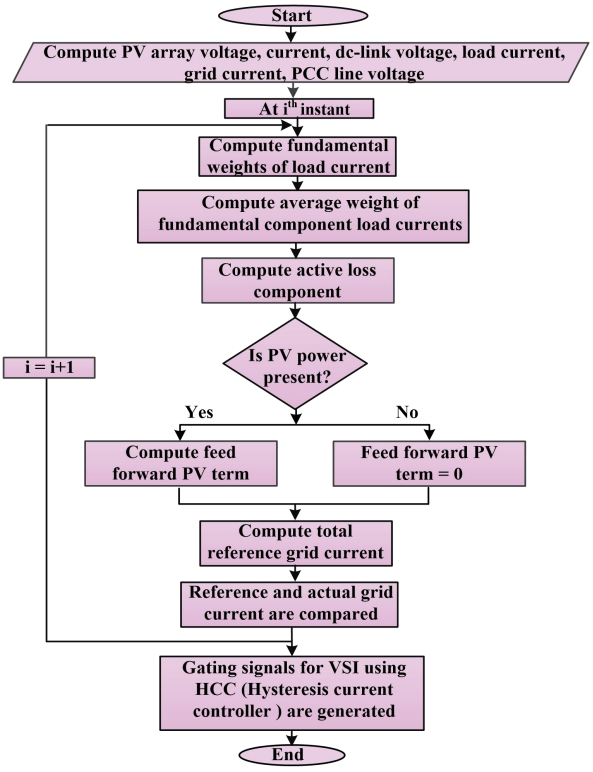


Fig. 3. Flowchart of the control in the grid coupled solar PV generating system.

IV. MATLAB/SIMULINK RESULTS

MATLAB/Simulink results are shown in this section for the performance evaluation of grid coupled solar PV generating system. The efficacy of the control scheme has been tested for different input and output conditions. Further, the abnormal grid conditions are considered to test the performance. Solar PV generating system is of 11.78kW, 3ϕ bridge rectifier with R=80ohm, L=20mH is considered as non-linear load. Three phase waveforms of grid voltage (V_g), grid current (I_g), load current (I_L), source current (I_s), dc link voltage (V_{dc}), average weight of fundamental load currents (I_{mean}), unit template and total reference grid current (I_{total}) are shown in the subsequent sections.

A. System performance under steady state

Under steady state condition, the VSI assists in maintaining the 3ϕ grid current to be balanced and sinusoidal. Also, the VSI balances the active power flow and maintain harmonics distortion in grid current within the IEEE-519 norms. Fig. 4 (a) depicts the waveforms of current and power of grid, load and VSI. Under steady state condition bridge rectifier with RL load (R=80Ω,20mH) has been taken as a nonlinear load till 0.3seconds and after 0.3 seconds 2kVA, 0.7 p.f. lagging load (1.4 kW and 1.4kVAR) has been increased in the system. However, till 0.3 seconds, nominal power(11kW) of the PV is supplying to the nonlinear load(3kW) and reserve power 8kW is fed to the grid. After 0.3 second total load has been increased to 4.4kW and 1.4 kVAR. System is still maintained the power balance. It can be noted that despite the nonlinear load, grid current is balanced and sinusoidal. After feeding the load, surplus active power of the source is supplied to grid as voltage and current of grid are in phase opposition. For unity power factor (UPF), reactive power of grid is zero. Voltage across dc link (V_{dc}) is maintained at 800V. At last, the harmonics spectrum of load current and grid current of phase 'a' are shown in Fig. 4 (b) & (c).

B. System performance under dynamic state (Insolation variation and load variation)

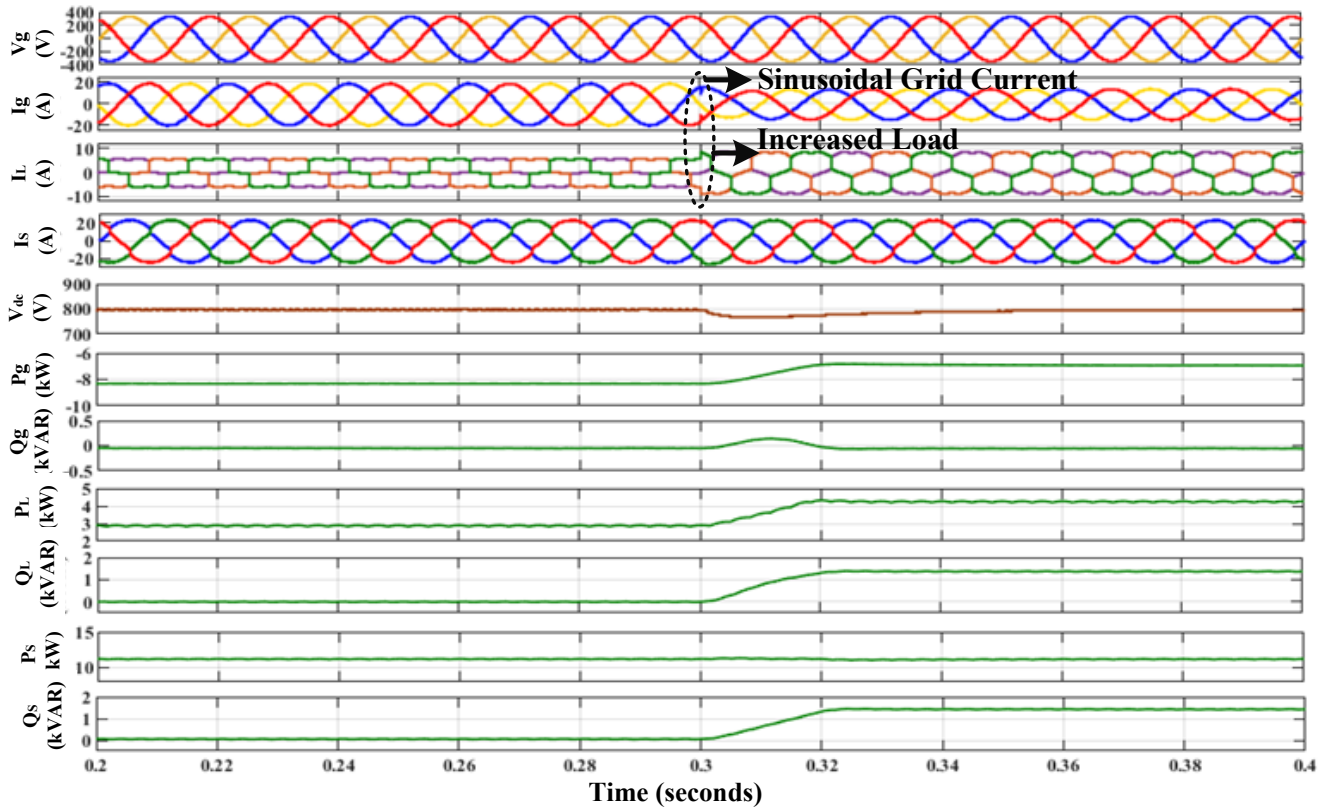
Under dynamic condition, the performance of the proposed control scheme has been illustrated in Fig. 5 at insolation and load variation. At 2.5s, the insolation level is dropped from 1000W/m^2 to 500W/m^2 that leads to decrease in PV power. With the decrease in the irradiation level, the active power generated from the PV decreases from 11.78kW to 5.2 kW. Hence, deficit power is demanded from the grid. This results in decrease in source current (I_s) and grid current (I_g) also reduces as PV is supplying power to grid. At 3.5s, new nonlinear load of R=50ohm, L=60mH has been added to the existing system. Now, the deficit load demand is fulfilled by grid as it can be seen that voltage and current of grid are not in phase opposition. Voltage across dc (V_{dc}) bus is maintained at 800V. The variation in I_{wmean} relies on the fundamental load current mean value as computed by equation (11). The crossing of zero value of I_{total} depicts the phase reversal of grid current.

C. System performance under abnormal grid condition (Voltage swell and voltage sag)

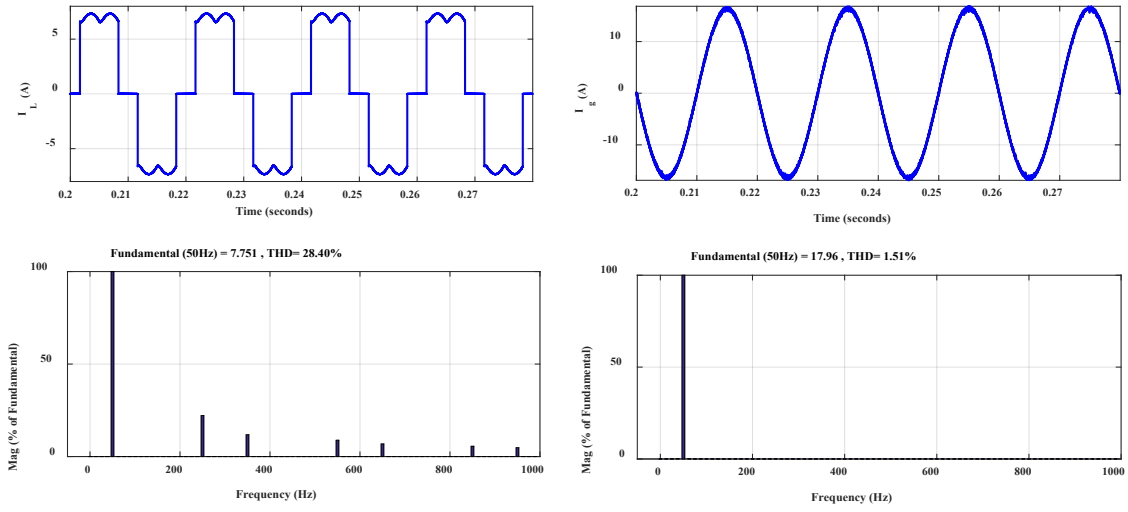
Under this section, the performance of the system has been analyzed at dynamic condition of voltage swell from 0.25 to 0.35s and sag from 0.35 to 0.4s is given in grid voltage as shown in Fig. 6. Voltage sag of 81V (P-P) at t= 0.25 to 0.35sec, decreases the current across the load. Alike, the voltage swell of 82V (P-P) at t= 0.35 to 0.4 sec, increases current across the load. Despite abnormal grid condition, grid currents are found to be balanced and sinusoidal. Voltage across dc (V_{dc}) bus is maintained at 800V. The variation in I_{wmean} relies on the fundamental load current mean value as computed by equation (11). The changes held in I_{total} is governed by equation (15) that depends upon $I_{wmean}, I_{loss}, I_{PV-ff}$ i.e., average weight of fundamental load currents, active loss component and feed-forward PV current respectively.

D. System performance under unbalanced condition (load unbalancing and grid unbalancing)

Under this section, the performance of the system has been analyzed at dynamic condition of load unbalancing from 0.25 to 0.35s. Load unbalancing from 0.35 to 0.4s is created by removing out one phase of load current. Unbalancing grid voltage is created by increasing the grid voltage of phase 'b' by 10% and reducing the grid voltage of phase 'c' by 10% from 0.35 to 0.4s as shown in Fig. 7. Grid current are found to balanced and sinusoidal under both the conditions. Voltage across dc (V_{dc}) bus is maintained at 800V. The variation in I_{wmean} relies on the fundamental load current mean value as computed by equation (11). The changes held in I_{total} is governed by equation (15) that depends upon I_{wmean}, I_{loss}, I_{PV-ff} i.e., average weight of fundamental load currents, active loss component and feed-forward PV current respectively.



(a)



(b) (c)

Fig. 4. System performance (a) Under steady state (b) Harmonic spectrum of load current-28.40% (c) Harmonic spectrum of grid current-1.51%

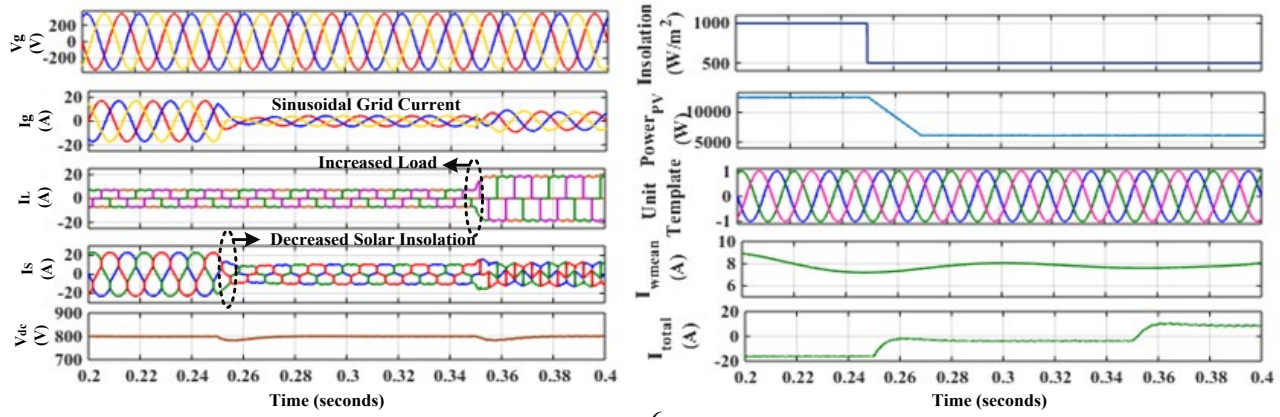


Fig. 5. System performance under dynamic state (Insolation variation and load variation)

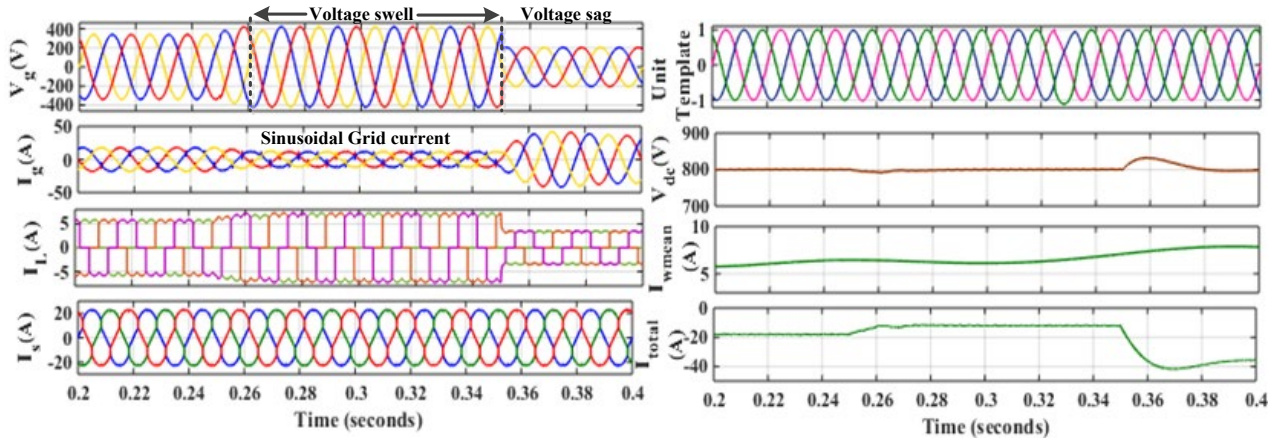


Fig. 6. System performance under abnormal grid condition (Voltage swell and voltage sag)

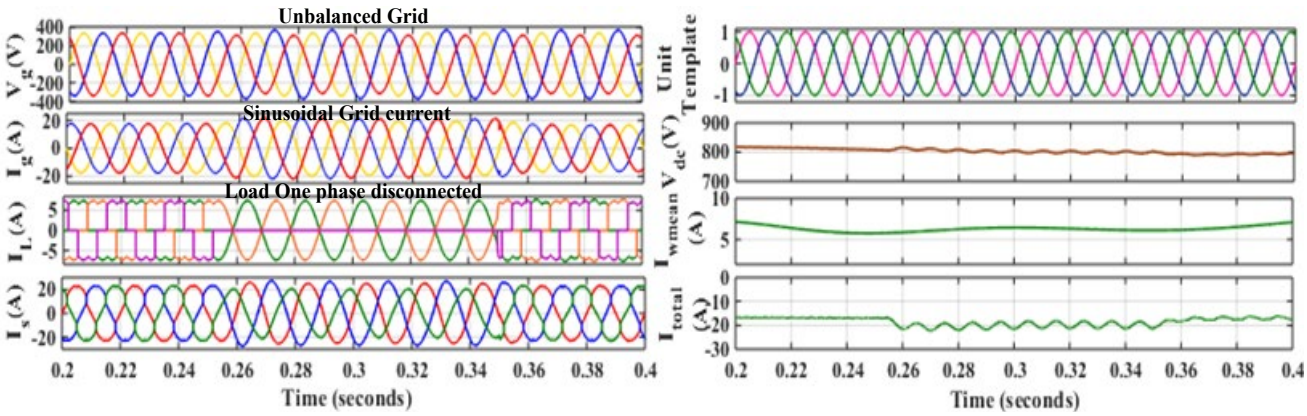


Fig. 7. System performance under unbalanced condition (load unbalancing and grid unbalancing)

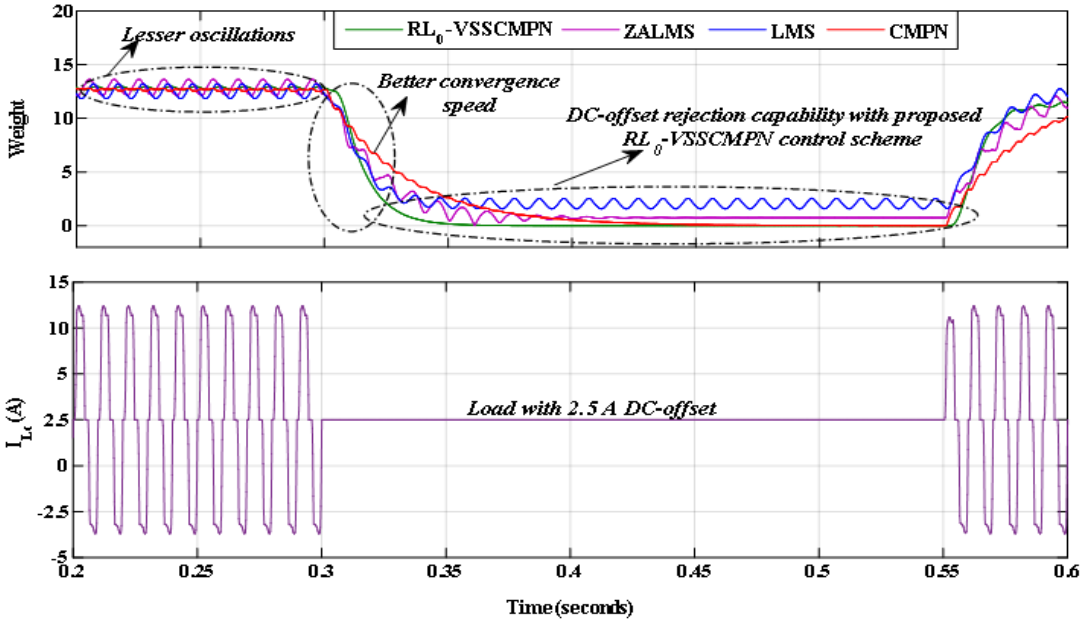


Fig. 8. Comparison of RL_0-VSSCMPN with other adaptive algorithms

Table 1. Comparison of RL_0-VSSCMPN with other adaptive algorithms

Algorithms	Oscillations	Convergence Speed	DC-offset rejection capability
RL_0-VSSCMPN	No	Fast	Complete
ZALMS	Less	Medium	Less
LMS	High	Slow	No
CMPN	High	Very Slow	Complete

E. Performance comparison of RL$_0$-VSSCMPN with other adaptive algorithms

In Fig.8 (a), comparison of weight convergence of proposed RL$_0$-VSSCMPN control scheme with other adaptive algorithms has been performed. It has been observed that the weight convergence with RL$_0$-VSSCMPN control scheme has less oscillation as well as has faster convergence speed as compared to other algorithms. Also, from Fig. 8 (a) it can be seen that proposed RL$_0$-VSSCMPN control scheme has zero steady state perturbations.

Further, an offset of 2.5A in load current has been created as shown in Fig. 8 (b). But it can be seen from Fig. 8 (a) that proposed RL$_0$-VSSCMPN control scheme has the DC-offset rejection capability.

F. Performance under parameters used in hardware

Solar PV generating system is of 1.8kW, 3ϕ bridge rectifier with R=50ohm, L=20mH is considered as non-linear load. Three phase waveforms of grid voltage (V_g), grid current (I_g), load current (I_L), source current (I_s), dc link voltage (V_{dc}) is shown in the Fig. 9, in which load has been increased at 0.3 second from 50 ohm to 30 ohm and waveforms are shown. Grid current are found to balanced and sinusoidal under both the conditions. Voltage across dc (V_{dc}) bus is maintained at 200V. THD is 1.9% which is within the limit as per IEEE standards.

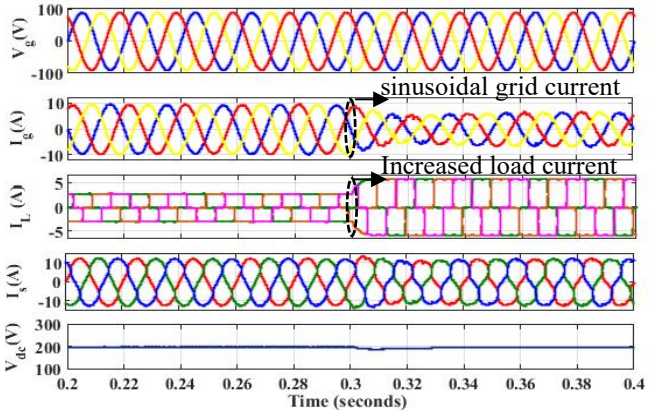


Fig. 9. System performance under parameters used in hardware

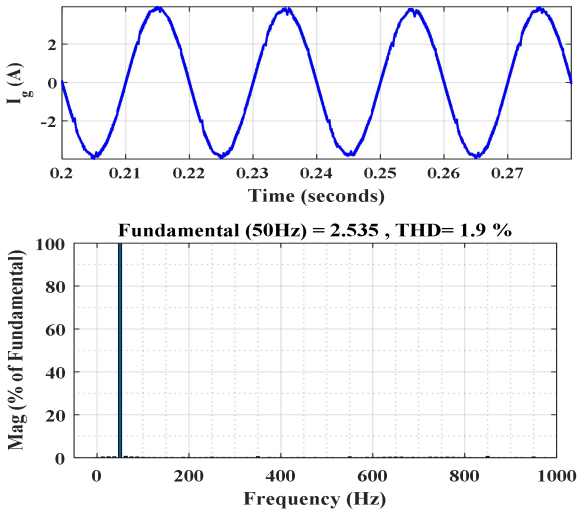


Fig. 10. THD window

V. Hardware Results

The efficacy of the proposed RL$_0$-VSSCMPN algorithm has been tested with hardware prototype developed in the laboratory acting as a distribution static compensator (DSTATCOM) shown in Fig. 11. Prototype system consists of VSI that is tied in parallel to the grid at the PCC. The voltage is sensed by voltage sensor and current is sensed using sensor based on Hall effect. FLUKE power analyzer has been used to measure response of the system. MicroLab box (dSPACE 1202) is used as controller. System parameters are given in appendix.

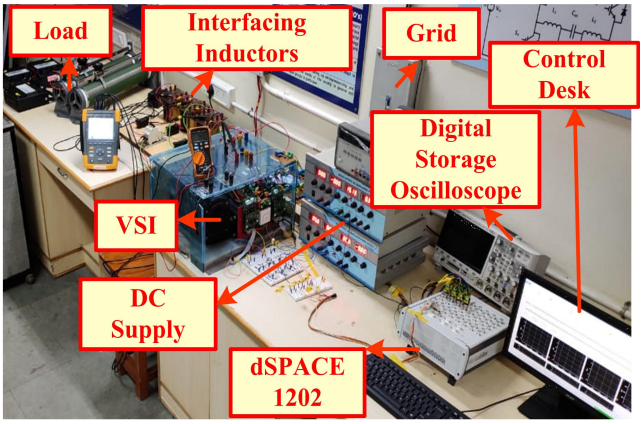


Fig. 11. Hardware Prototype

A. System performance under steady state

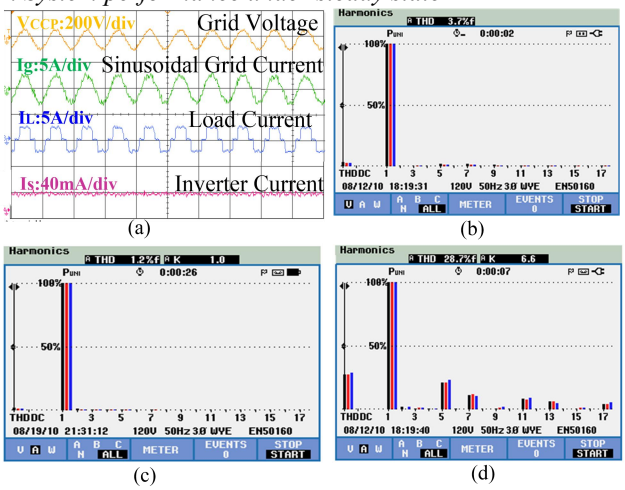


Fig. 12. System performance under steady state

Fig. 12 shows the steady state performance of the system with nonlinear load (3ϕ bridge rectifier). Fig. 12 (a) depicts the voltage waveform of common coupling point, current waveforms of grid, load and inverter.

In Fig. 12 (b)-(d), total harmonics distortion (THD) level of CCP voltage, grid and load current are shown. The THD in grid current was 1.2% only and it is under IEEE norms whereas load current has 28.7% THD. The CCP voltage THD is 3.7%.

B. System performance under grid unbalance

System performance under grid unbalance condition has been tested and experimental results is shown in Fig. 13. Under unbalancing of 16V(P-P) in phase 'a' of grid voltage, performance of proposed control are shown in the result. It can

be seen from the results that under unbalance grid condition, proposed control efficiently maintain the grid current to be sinusoidal.

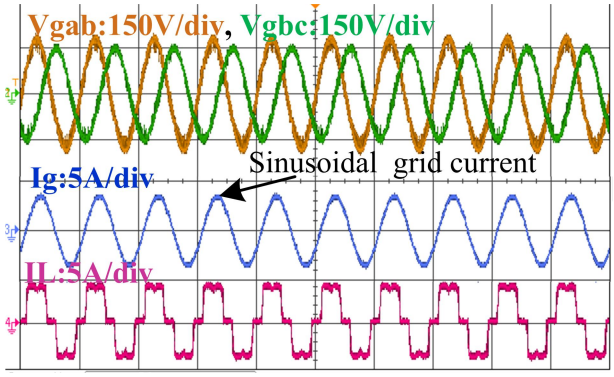


Fig. 13. System response under unbalanced grid voltages

C. System performance under load varying condition

In Fig. 14, efficacy of the proposed controller has been examined under varying non-linear load. At the instant marked, the non-linear load had been added to the existing system. A 3-ϕ bridge rectifier (R=50Ω)) has changed to (R=30Ω) at the marked instant. The grid current is sinusoidal. Furthermore, it is seen that grid current magnitude is increased to meet the increased load demand.

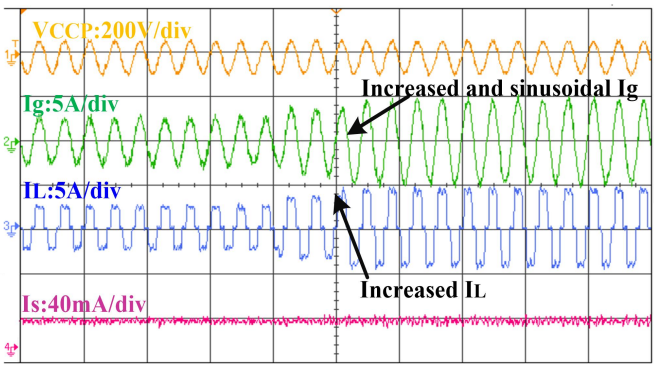


Fig. 14. System response under load varying condition

D. System performance under voltage sag and voltage swell

In Fig. 15 performance of the proposed controller under dynamic condition i.e., voltage sag and swell in the system is shown. Voltage sag of 100V at instant indicated, decreases the voltage across the load current and similarly, voltage swell of 150V at instant indicated, increases the voltage across the load current. The grid current is still observed to be sinusoidal in both the condition.

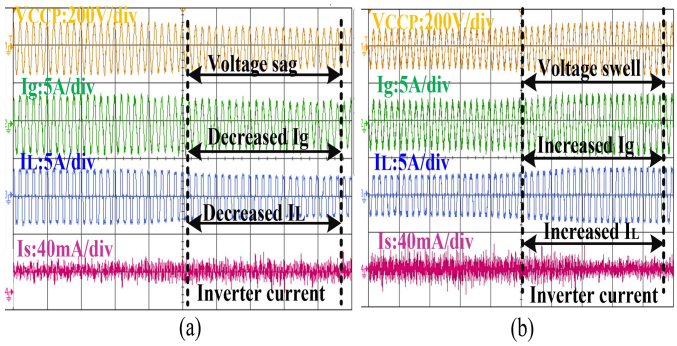


Fig. 15. System performance under voltage sag and voltage swell

E. System performance under unbalanced load condition

Fig. 16 shows the dynamic performance under unbalanced load condition. Unbalancing is created by removing out phase 'b' of load. It is seen that under unbalanced load condition, the grid current is still sinusoidal. Voltage across DC bus is maintained at 200V. Fig. 16 (a) is compared with (b) in terms of I_{total} as obtained from equation (15). Proposed RL_0-VSSCMPN is compared with conventional LMS and it is seen that with RL_0-VSSCMPN the oscillations are minimized and has faster convergence speed.

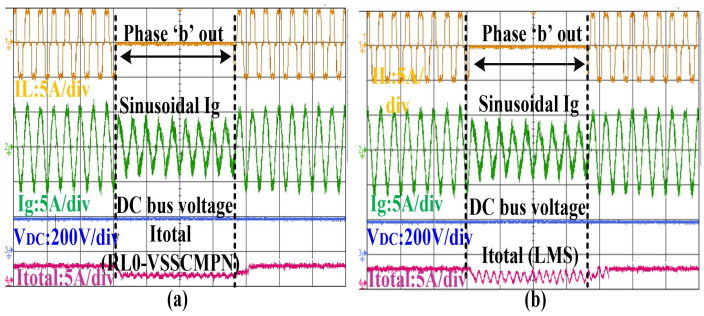


Fig. 16. System performance under unbalanced load condition

VI. Conclusion

A novel control scheme reweighted L_0 norm variable step size continuous mixed p-norm (RL_0-VSSCMPN) for maintaining balanced and sinusoidal grid current in grid coupled SPVGS was investigated under different scenarios such as steady state and dynamic load condition, abnormal and unbalanced grid and with dc offset. The adaptability of the developed algorithm is justified by comparing the performance with other conventional algorithms. The fundamental load component with proposed control RL_0-VSSCMPN has less steady state oscillations and fast convergence speed. The obtained THD in grid current with non-linear load are also within IEEE norms Moreover, from simulation results, it is seen that VSI supplies the reactive power demand of the load, maintained DC bus voltage, minimizes harmonics and load balancing. The proposed algorithm is also tested on a prototype hardware developed in the laboratory of Delhi Technological University. The results obtained from simulation and hardware are found to be satisfactory under steady state as well as dynamic scenarios.

Appendix

A) Simulation Parameters : 3-phase AC mains: 415V (Line-line), 50Hz; Interfacing Inductors L_f =5mH; 3-phase bridge rectifier non linear load R= 80Ω, L=20mH; dc link voltage V_{dcref} = 800V; dc PI gain K_{pd} = 0.5, K_{id} = 5, $\gamma = 0.02$, $f_\beta = \pm 1, \mu = 0.0003$.

B) Experimantal Parameters: 3-phase AC mains:110V (Line-line), 50Hz; Interfacing Inductors L_f =5mH; 3-phase bridge rectifier non linear load R= 50Ω; dc link voltage V_{dcref} = 200V; dc PI gain K_{pd} = 0.5, K_{id} = 0.01 $\gamma = 0.02$, $f_\beta = \pm 1, \mu = 0.0003$

References

[1] J. J. Nedumgatt, K. B. Jayakrishnan, S. Umashankar, D. Vijayakumar, and D. P. Kothari, "Perturb and observe MPPT algorithm for solar PV systems-modeling and simulation," *Proc. - 2011 Annu. IEEE India Conf. Eng. Sustain. Solut. INDICON-2011*, vol. 19, no. 1, pp. 1–6, 2011.

[2] M. Lokanadham and K. V. Bhaskar, "Incremental Conductance Based Maximum Power Point Tracking (MPPT) for Photovoltaic

System," *Int. J. Eng. Res. and App.*, vol. 2, no. 2, pp. 1420–1424, 2012.

[3] B. Yang *et al.*, "Perturbation observer based fractional-order sliding-mode controller for MPPT of grid-connected PV inverters: Design and real-time implementation," *Control Eng. Pract.*, vol. 79, no. June, pp. 105–125, 2018.

[4] T. Esram and P. L. Chapman, "Comparison of Photovoltaic Array Maximum Power Point Tracking Techniques," *IEEE Trans. Energy Convers.*, vol. 22, no. 2, pp. 439–449, 2007.

[5] R. K. Kharb, S. L. Shimi, S. Chatterji, and M. F. Ansari, "Modeling of solar PV module and maximum power point tracking using ANFIS," *Renew. Sustain. Energy Rev.*, vol. 33, pp. 602–612, 2014.

[6] A. Menadi, S. Abdeddaim, A. Betka, and M. T. Benchouia, "Real Time Implementation of A Fuzzy Logic Based Mppt Controller for Grid Connected Photovoltaic System," *Int. Jou. of Renewable Energy. Res.*,vol. 5, no. 1, 2015.

[7] N. Patel, N. Gupta, and B. C. Babu, "Design, development, and implementation of grid-connected solar photovoltaic power conversion system," *Energy Sources, Part A Recover. Util. Environ. Eff.*, pp. 1–20, 2019.

[8] K. Arulkumar, K. Palanisamy, and D. Vijayakumar, "Recent advances and control techniques in grid connected Pv system - A review," *Int. J. Renew. Energy Res.*, vol. 6, no. 3, pp. 1037–1049, 2016.

[9] B. Singh, C. Jain, S. Goel, R. Gogia, and U. Subramaniam, "A Sustainable Solar Photovoltaic Energy System Interfaced with Grid-Tied Voltage Source Converter for Power Quality Improvement," *Electr. Power Components Syst.*, vol. 45, no. 2, pp. 171–183, 2017.

[10] R. Abbassi, "SOGI-based Flexible Grid Connection of PV Power Three Phase Converters under Non-ideal Grid Conditions," *Eng. Technol. Appl. Sci. Res.*, vol. 10, no. 1, pp. 5195–5200, 2020.

[11] H. M. Hasanien, "An Adaptive Control Strategy for Low Voltage Ride Through Capability Enhancement of Grid-Connected Photovoltaic Power Plants," *IEEE Trans. Power Syst.*, vol. 31, no. 4, pp. 3230–3237, 2016.

[12] G. Gui, A. Mehbodniya, and F. Adachi, "Least mean square/fourth algorithm for adaptive sparse channel estimation," *IEEE Int. Symp. Pers. Indoor Mob. Radio Commun. PIMRC*, pp. 296–300, 2013.

[13] M. Karimi-ghartemani, "Linear and Pseudolinear Enhanced Phased-Locked," *IEEE Trans. Ind. Electron.*, vol. 61, no. 3, pp. 1464–1474, 2014.

[14] F. Wu, L. Zhang, and J. Duan, "A new two-phase stationary-frame-based enhanced PLL for three-phase grid synchronization," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 62, no. 3, pp. 251–255, 2015.

[15] S. Gude and C. C. Chu, "Dynamic Performance Improvement of Multiple Delayed Signal Cancelation Filters Based Three-Phase Enhanced-PLL," *IEEE Trans. Ind. Appl.*, vol. 54, no. 5, pp. 5293–5305, 2018.

[16] M. Karimi-Ghartemani, S. A. Khajehoddin, P. K. Jain, A. Bakhshai, and M. Mojiri, "Addressing DC component in pll and notch filter algorithms," *IEEE Trans. Power Electron.*, vol. 27, no. 1, pp. 78–86, 2012.

[17] A. Mortezaei, C. Lute, M. G. Simões, F. P. Marafão, and A. Boglia, "PQ, DQ and CPT control methods for shunt active compensators - A comparative study," *2014 IEEE Energy Convers. Congr. Expo. ECCE 2014*, pp. 2994–3001, 2014.

[18] S. Nanda and P. K. Dash, "A Gauss-Newton ADALINE for dynamic phasor estimation of power signals and its FPGA implementation," *IEEE Trans. Instrum. Meas.*, vol. 67, no. 1, pp. 45–56, 2018.

[19] N. Gupta and R. Garg, "Tuning of asymmetrical fuzzy logic control algorithm for SPV system connected to grid," *Int. J. Hydrogen Energy*, vol. 42, no. 26, pp. 16375–16385, 2017.

[20] F. Lin, K. Lu, T. Ke, B. Yang and Y. Chang, "Reactive Power Control of Three-Phase Grid-Connected PV System During Grid Faults Fuzzy Neural Network Control," IEEE Transactions on Industrial Electronics, vol. 62, no. 9, pp. 5516-5528, Sept. 2015.

[21] Y. He, Y. Zou, F. Liu, and H. Li, "Research on an improved adaptive algorithm applied to active power filters," *Proc. IEEE Power Eng. Soc. Transm. Distrib. Conf.*, pp. 672–677, 2006.

[22] A. K. Pradhan, A. Routray, and A. Basak, "Power system frequency estimation using least mean square technique," *IEEE Trans. Power Deliv.*, vol. 20, no. 3, pp. 1812–1816, 2005.

[23] B. Sun, Z. Teng, Q. Hu, S. Tang, W. Qiu, and H. Lin, "A Novel LMS-Based SANC for Conveyor Belt-Type Checkweigher," *IEEE Trans. Instrum. Meas.*, vol. 70, no. 30, pp. 1–10, 2021.

[24] M. Badoni, A. Singh, and B. Singh, "Comparative Performance of Wiener Filter and Adaptive Least Mean Square-Based Control for Power Quality Improvement," *IEEE Trans. Ind. Electron.*, vol. 63, no. 5, pp. 3028–3037, 2016.

[25] M. Mangaraj and A. K. Panda, "Modelling and simulation of KHLMS algorithm-based DSTATCOM," IET Power Electron., vol. 12, no. 9, pp. 2304–2311, 2019.

[26] R. K. Agarwal, I. Hussain, and B. Singh, "Implementation of LLMF Control Algorithm for Three-Phase Grid-Tied SPV-DSTATCOM System," *IEEE Trans. Ind. Electron.*, vol. 64, no. 9, pp. 7414–7424, 2017.

[27] A. Arora and A. Singh, "Design and analysis of functional link artificial neural network controller for shunt compensation," IET Gener. Transm. Distrib., vol. 13, no. 11, pp. 2208–2215, 2019.

[28] S. Kewat and B. Singh, "Improved Reweighted Zero-Attracting Quaternion-Valued LMS Algorithm for Islanded Distributed Generation System at IM Load," *IEEE Transaction Industrial Elec.*,vol. 67, no. 5, pp. 3705–3716, 2020.

[29] J. Chambers and A. Avlonitis, "A robust mixed-norm adaptive filter algorithm," *IEEE Signal Process. Lett.*, vol. 4, no. 2, pp. 46–48, 1997.

[30] L. Shi, H. Zhao, and Y. Zakharov, "Generalized Variable Step Size Continuous Mixed p-Norm Adaptive Filtering Algorithm," *IEEE Trans. Circuits Syst. II Express Briefs*, vol. 66, no. 6, pp. 1078–1082, 2019.

[31] H. Zayyani, "Continuous mixed p-norm adaptive algorithm for system identification," *IEEE Signal Process. Lett.*, vol. 21, no. 9, pp. 1108–1110, 2014.

[32] A. Ranjan, Seema, and B. Singh, "Reweighted L1 Norm penalized Least Mean Square Fourth Algorithm based Solar Grid Interfaced System for Alleviating Power Quality Problems," *2020 IEEE Int. Conf. Power Electron. Smart Grid Renew. Energy, PESGRE 2020*, vol. 56, no. 5, pp. 5352–5362, 2020.

[33] K. Tiwari, Seema, and B. Singh, "CMPN Adaptive Control Algorithm for Double Stage PV-Grid Connected System," *2020 IEEE Int. Conf. Power Electron. Smart Grid Renew. Energy, PESGRE 2020*, pp. 2–7, 2020.

[34] Z. Li, "Continuous Mixed p -Norm Adaptive Algorithm with Reweighted L0-Norm Constraint," *IEEE Int. Conf. on Elect. Inst. & inf sys.*no. 1, pp. 1–5.

BIOGRAPHY



Pallavi Verma, member of IEEE. Received her master from NITTTR and Ph.D. from Punjab University respectively. Her area of interest is modeling of operation and control, smart grid and renewable energy.



Avdhesh Kumar, member of IEEE. Received the master degree in electrical engineering from IIT BHU Varanasi. He is currently pursuing Ph.D. degree with the department of electrical engineering at the Delhi technological university (Delhi), India.



Rachana Garg, Senior member of IEEE. Received the B.E. and M.E. degree in 1986 and 1989 respectively from NIT, Bhopal, India. She has obtained her Ph.D. in Electrical Engineering from Delhi University, India in 2009. Presently, she is working as Professor in DTU, Delhi, India. Her area of interest is modeling of transmission lines, power system operation and control, smart grid and



Priya Mahajan member of IEEE. Received the B.E. and M.E. degree in 1996 and 1998 from Thapar University and Punjab Engineering College, India respectively. She has obtained her Ph.D. in Electrical Engineering from Delhi University, India in 2015. Presently, she is working as Professor in Delhi Technological University, Delhi, India. Her area of interest is Railway Electric Traction System, Power System, Electrical Machines.

Applied Intelligence
https://doi.org/10.1007/s10489-022-03642-w

Semantic segmentation in medical images through transfused convolution and transformer networks

Tashvik Dhamija[1] · Anunay Gupta[2] · Shreyansh Gupta[3] · Anjum[4] · Rahul Katarya[4] · Ghanshyam Singh[5]

Accepted: 15 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Recent decades have witnessed rapid development in the field of medical image segmentation. Deep learning-based fully convolution neural networks have played a significant role in the development of automated medical image segmentation models. Though immensely effective, such networks only take into account localized features and are unable to capitalize on the global context of medical image. In this paper, two deep learning based models have been proposed namely USegTransformer-P and USegTransformer-S. The proposed models capitalize upon local features and global features by amalgamating the transformer-based encoders and convolution-based encoders to segment medical images with high precision. Both the proposed models deliver promising results, performing better than the previous state of the art models in various segmentation tasks such as Brain tumor, Lung nodules, Skin lesion and Nuclei segmentation. The authors believe that the ability of USegTransformer-P and USegTransformer-S to perform segmentation with high precision could remarkably benefit medical practitioners and radiologists around the world.

Keywords COVID-19 · Deep learning · Image-segmentation · Transformer · U-net

1 Introduction

SINCE the dawn of the 21st Healthcare industry has started adopting technology rapidly. One area in which this adoption of technology has given remarkable results is the area of "Medical Imaging". Medical imaging refers to techniques and processes used to mimic the state of organs via images CT-Scan, X-rays and MRI [1]. Until the past decade, the advancements in medical imaging were more focused on optimizing and improving the process of creating organ images and enhancing the quality of medical images. While the process of inferring information was left untouched and immensely dependent on the availability of experts and trained professionals. The lack of innovation in information inferring processes has led to burdened healthcare infrastructures in many countries. The only way to reduce this burden is to automate certain aspects of inferring information from medical images. One domain that has a lot of scope automation is segmentation.

✉ Rahul Katarya
rahuldtu@gmail.com

Tashvik Dhamija
tashvik369@gmail.com

Anunay Gupta
aganunay@outlook.com

Shreyansh Gupta
shreyansh.gupta1299@gmail.com

Anjum
anjum_2792@yahoo.com

Ghanshyam Singh
ghanshyams@uj.ac.za

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

[2] Department of Electrical Engineering, Delhi Technological University, New Delhi, India

[3] Department of Civil Engineering, Delhi Technological University, New Delhi, India

[4] Department of Computer Science Engineering, Delhi Technological University, New Delhi, India

[5] Department of Electrical and Electronic Engineering Science, University of Johannesburg, Johannesburg, South Africa

Published online: 25 April 2022

Segmentation in medical imaging deals with labelling each pixel on the image with a class which is known as Semantic segmentation. Semantic Segmentation has a plethora of applications in the healthcare industry. Some of the areas where segmentation has been used are - detection of lung nodules from CT-Scans of lungs [2], Detection of Brain tumor [3], segmentation of skin lesion [4], Polyp detection [5] (Polyps are irregular tissue growth on body), Liver segmentation [6], Nuclei segmentation [7], etc. A lot of research has been put into developing segmentation models and algorithms using multiple toolboxes. A lot of research has been put into developing segmentation models and algorithms using multiple toolboxes. After going through literature, we found that, until now the problem of medical image segmentation has been tackled with algorithms that can be grouped into 4 broad categories namely, 1) machine learning-based algorithms, 2) early convolution neural networks (CNN), 3) optimized CNN, and 4) feature extracting convolution networks. Each category of algorithms is presented as follows [8-10].

1.1 Machine learning based algorithms

Machine learning methods in the medical image segmentation are broadly classified into two categories such as 1) supervised learning and 2) unsupervised learning. The supervised methods employed artificial neural networks [11]. However, the unsupervised methods consisted of extensive use of K-mean algorithm [12], Hard C-mean algorithms [13] and Fuzzy C-means algorithms [14]. The common problem among these techniques is that such techniques majorly rely on predefined features or some structure. The reliance on predefined features reduces the generalizing ability and robustness of the techniques.

1.2 Early convolutional networks

With the significant development in the deep learning, the task of medical image segmentation has tried to solved using Convolutional Neural Networks (CNN) [15] models. Early on, the patching-based CNN models are used over primitive algorithms as reported in [16]. Then, a ground-breaking improvement has been observed with the invent of the UNet [17]. It builds upon the concept of using a fully convolutional network (FCN) and it consists contracting segment that encodes the features and an expansive segment that generates the mask from the encoded features. This architecture is capable of learning high-resolution features to provide more precise outputs. The authors also introduced the use of data augmentation during training to aid the training of models on smaller datasets and to make them robust. The limitation of the UNet architecture is that they use simple skip connections which lack the ability to transfer spatial features across the encoder and decoder.

1.3 Optimized convolutional networks

In order to overcome the limitation of simple skip connections in the Unet, multiple architectures with complex skip connections are introduced. in [18], the authors have proposed an architecture named, MultiResUNet. This model replaces pairs of standard convolution layers present in the vanilla UNet with inception like MultiRes block to restore features learned at various scales while maintaining memory efficiency. The MultiRes block also consists of 1×1 convolutions to learn spatial information, owing to such additions MultiResUnet passes baselines set by Unet on multiple datasets. Further, the authors in [19] introduced DC-UNet that builds upon the MultiResUNet and aimed to improve the performance of image segmentation algorithms on tricky dataset. Further, the authors replace the simple residual connection in the MultiRes block with a sequence of 3×3 convolutions to increase the ability of the model to learn better spatial features. This architecture can learn features of different scales, reduce the semantic gap and learn spatial features. The DC-UNet is able to perform considerable well on infrared (IR) breast dataset, IEEE-ISBI dataset, and CVC_Clinic DB. Another approach as reported in [20] is taken to introduce efficient skip connections in the original UNet. The authors proposed a novel architecture, the UNET++ for semantic and instance image segmentation. The architecture differs from the original UNet architecture proposed in [17] in terms of optimized skip connections and connections between adjacent nodes. Thus, such modifications equip UNET++ to have considerable advantages over the traditional UNet architecture like always having the optimal depth and efficient fusion of features in the decoder. Further UNET++ achieves decent results on six problems in the domain of medical image segmentation, namely, Electron Microscopy (EM), Cell segmentation, Nuclei segmentation, Brain tumor segmentation, Liver Segmentation, and Lung Nodule Segmentation. The third way to improve the UNet by making the skip connections more effective has explored in [21] where the authors of BCDU-NET make use of Bi-ConvLSTM in the skip connections, which assist in relaying semantic information between the corresponding layers. The effectiveness of this result is proved by achieving state-of-the-art results on three benchmark datasets, namely, Drive Dataset, ISIC 2018 Dataset, and Lung Nodule Analysis (LUNA) dataset.

1.4 Feature extraction convolutional networks

In order to further enhance the performance of image segmentation algorithms, multiple newer architectures which focus on improving the feature extraction power of the UNet encoder have been introduced. In [22], the authors have proposed DoubleU-Net model for medical image segmentation which is based on the idea of using two UNets wherein the first UNet

uses a VGG-19 encoder pre-trained on ImageNet and the second UNet takes in the multiplication of the original image and the output of the first UNet as input. The DoubleU-Net performs well on datasets such as CVC-Clinic DB, ISIC 2018, and 2018 Data science Bowl challenge datasets. Further, a different approach has been taken by the authors of a novel segmentation architecture DRINet [23], which is based on DenseNet and Inception-ResNet. The dense blocks in the architecture of DRINet are used in the encoder part and residual inception blocks, along with unpooling from the decoder of the DRInet. Thus, such blocks are used to replace standard convolution blocks to learn distinctive features in cases of similar shape, intensity, location, and size. The DRINet performs well on CSF, CT, and multi-organ datasets. In [24], the authors have proposed another Novel architecture for 2D medical image segmentation, namely, CE-NET which has 3 major components such as Feature Encoder, Content extractor, and decoder in contrast to encoders, and decoders seen in typical UNets. The use of a context extractor that performs Atrios convolutions help to make the filters act on global features and thus provides the model more context. The CE-NET achieves excellent results in Optic disc image segmentation, Retinal Vessel Detection, Lung segmentation, and Cell contour segmentation.

A potential research challenge in the aforementioned architectures is that these models are only capable of extracting local context and unable to take into account global features, and long-range dependencies in the image data. Despite the heavy optimization in the extraction and usage of local features, the improvements in models' performance are significantly low. We hypothesize that this is due to the lack of consideration of global features and intending to overcome this research gap and fully capitalize on global features in segmenting medical images. We propose the USegTransformer-P and USegTransformer-S approach which leverage both local features from the full convolution networks and long-term dependencies obtained by transformers. We believe that such a combination will yield state-of-the-art results. The novelty and salient features of this paper is summarized as follows:

- We have developed a novel medical image segmentation models, USegTransformer-P and USegTransformer-S, which utilizes both the local and global features of the image using an amalgamation of Transformers based encoders and UNet based encoders.
- We have developed a fully convolutional ensemble decoder for transfusing both the local and global features obtained from a fully convolutional network and transformer network respectively, from an image in a learnable manner.
- We have analyzed the performance of sequential and parallel stacking of transformer-based encoder-decoder and convolution-based encoder-decoder in order to achieve a better configuration of the proposed models.
- We have proved the efficacy of the proposed models by comparing them with reported medical image segmentation models on benchmark datasets used in renowned competitions such as LUNA, ISIC-2018, and Kaggle Data science bowl.

The paper is further divided into 4 sections, namely, the methodology, results and experimentation, discussion, and conclusion and future scope. In the methodology section, we explain every intricate detail about our proposed models, USegTransformer-P and USegTransformer-S. In the result section, we explain and analyze the performance of our proposed model with the help of metrics and visualizations. In the discussion section we probe into what are qualitative and quotative effects of the use of global features in segmentation algorithms. In the final section that is conclusion and future work; we present our findings as well as discuss the direction future works that can be done in the field in the automated medical segmentation.

2 Proposed architecture

The UNet-based model extracts localized features in the high-level representations such as the CNN models. On the other hand, transformer-based models like ViT extract global context and long-range dependencies. In a task like an image segmentation, we would require both kinds of feature representations. Therefore, in this proposed system model, we come up with two different methods to use the best of both worlds. These methods take advantage of the different feature extraction abilities of a transformer model and CNN model. By doing this, the proposed system model is capable of understanding the local features as well as the global context. The first model, known as the USegTransformer-P shown in Fig. 2a, is a parallel model and the second model that is the USegTransformer-S as shown in Fig. 2b, is a sequential model inspired from the work reported in [22].

The proposed system models derive their name from the use of a transformer-based encoder-decoder architecture stacked with U-Net inspired encoder-decoder convolutional architecture to segment the bestowed image. The suffix of USegTransformer (-P and -S) signifies the type of stacking. For the given image $img \in R^{H \times W \times C}$ with a spatial resolution of $H \times W$, where H and W denote height and width of the conferred image and C channels. The main goal is to yield a segmentation mask $msk \in R^{H \times W}$. The architecture of USegTransformer-P and USegTransformer-S consists of two common parts, namely, the transformer-based encoder-decoder and an UNet inspired encoder-decoder convolutional block.

These components and the stacking details of the proposed models are discussed in the following subsections.

2.1 Transformer based image encoder-decoder

The vision Transformer models have been quite successfully used in both computer vision [25] and natural language processing [26] and have provided promising results in both fields. The primary edge that the transformers have over other techniques is efficiency in terms of computational resource usage and its efficacy in performing various tasks. This success of the vision transformers model can be attributed to parallelization, where, unlike in Recurrent Neural Networks (RNNs) and Long Short-Term Memory (LSTM), all time-steps can be passed in a single step making the model computationally efficient. Secondly, the transformers use a self-attention mechanism to gain global attention features from the data which attributes to the results that the transformers achieve.

2.1.1 Transformer encoder

This research study used a transformer-based self-attention architecture to encode the images into high-level features with a global context. We have first divided the image $\{img^i \in R^{H \times W \times C}\}$ into two-dimensional (2D) flattened patches $\left\{img^i_{pat} \in R^{P^2.C}\right\}$, where $i \in [1 \ldots N]$ and each patch is of size $P \times P$ with $N = \frac{W.H}{P^2}$ to make the image data sequential and time distributed as reported in [25]. Further, we have also added a linear layer called the projection layer to project these flattened patches into a L dimension vector where L is the length of image sequence and these patches are known as Projected Patches. To include the spatial information of the input image in the proposed model, we added these projected patches with a positional encoding matrix $\left\{E_{Pos} \in R^{N \times Embed_{Size}}\right\}$ where EmbedSize is given as $Embed_{Size} = L$. The positional encodings matrix can be developed in different ways however, we have kept the positional encoding matrix as a learnable parameter in this proposed research, i.e., the model will also update the elements of this positional embedding matrix during back-propagation. The positional encodings and the flattened patches are combined and mathematically presented as:

$$P_{embed} = \left[x^1_t E, x^2_t E, x^3_t E, \ldots, x^N_t E\right] + E_{Pos} \tag{1}$$

where E is the linear projection matrix such that $E \in R^{\left(P^2 C \times L\right)}$ and E_{Pos} represent the learnable position encoding. After the image's encoding is performed, we move on to apply the Multi-Headed-Self-Attentions (MHSA) and Position-wise Feed-Forward Networks (FFNs) layers. The transformer consists of H Multi-Head-Self-Attention (MHSA) and H Position-wise Feed-Forward Networks (FFNs) blocks. The Eqs. (2) and (3) explain the operations performed during the self-attention phase.

$$\ddot{z}_j = MHSA\left(LinNorm\left(z_j\right)\right) + z_{j-1} \tag{2}$$

$$z_j = FFN\left(LinNorm\left(\ddot{z}_j\right)\right) + \ddot{z}_{j-1} \tag{3}$$

where $LinNorm(\,)$ denote the Linear Layer Normalization and z_j is a feature tensor. The FFN block includes linear layers with the rectified linear unit (ReLU) activation function, and the MHSA contains L self-attention heads (SAs) connected in parallel. In each SA head, we cmputed the attention by mapping queries and key-value pairs to an output. The key (K), value (V), and query (Q) matrixes are formed by the encoded input matrices. Now, we compute the dot of the query matrix with all the key matrices and then scale each of them by $\sqrt{d_k}$ where d_k is the dimensions of the key matrix and query matrix and apply the SoftMax function to obtain the self-attention weight matrix, which then multiplied with the value matrix. The Eq. (4) below explains this procedure mathematically.

$$Attention(K, Q, V) = Softmax\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{4}$$

The outputs from all the attention heads are concatenated into a single matrix, which is represented as:

$$MHSA(z) = [h_1, h_2, h_3 \ldots, h_H]W_{MHSA} \tag{5}$$

where W_{MHSA} represents the weights of the learnable weight matrix of different SAs. The flow of input through the transformer encoder layer can also been understood by analysing Fig. 1.

2.1.2 Convolution based decoder

After obtaining the encoded images through the transformer-based encoder, we pass the encoded vector, which is first reshaped and rearranged to a 2D matrix $Img_{encode} \in R^{\ddot{w} \times \ddot{h} \times c}$ and then passed through a convolutional auto-encoder inspired decoder as reported in [27]. Here, we have used decoder convolutional up-sampling blocks. Each convolutional up-sampling block includes a Conv2D layer with a kernel size of 3×3 with a Batch-Normalization layer and ReLu activation with a bilinear up-sampling layer that has a scale factor of 2. The image from the bottleneck encoder is passed through each decoding step where at every step, we up-sample the image by a factor of 2. This process is continued until we achieve the

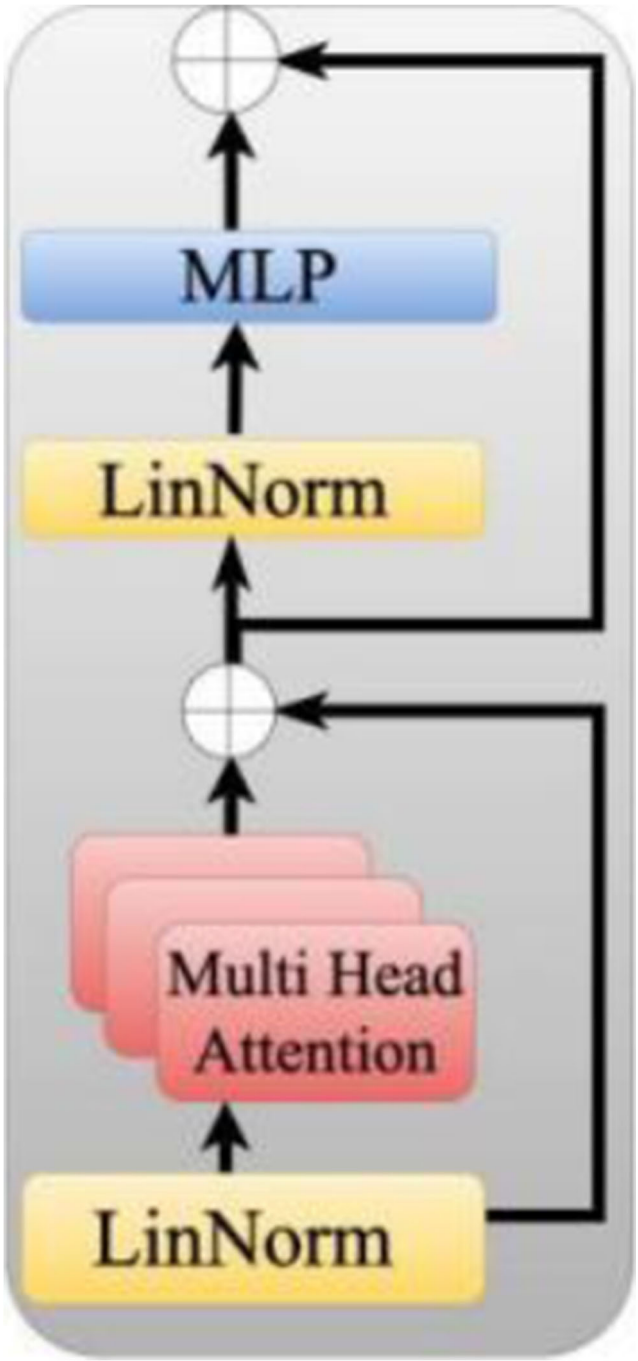


Fig. 1 Schematic of the transformer based image encoder

original image resolution. At the final stage, we use a convolution layer with 1 × 1 kernel size and with the sigmoid activation in the end.

2.2 U-net inspired encoder-decoder block

As pellucid from the above discussion, the transformer-based encoder provides a global high-level representation of features. Therefore, we need a fully convolutional network to yield localized high-level feature representation to identify subtly local regions in the masked image. Hence, we use a UNet-inspired fully convolutional encoder-decoder architecture for this purpose. The encoder block consists of a sequence of 2D convolution, batch normalization, and activation, repeated twice, followed by a pooling function to downscale. The decoder block consists of a sequence of linear or bilinear up-sampling followed by 2D convolution, batch normalization, and activation function, repeated twice. Additionally, there exists a stacking of encoder outputs to decoder inputs at the same dimension across the encoder and the decoder through skip connections. The flow of input image in a UNet can be mathematically expressed as:

$$R^{\epsilon H\times W\times 3}\Rightarrow R^{\epsilon h\times w\times c}\Rightarrow R^{\epsilon H\times W\times C} \tag{6}$$

In the experiments, all convolutions have a 3 × 3 kernel with padding = 1. The activation function uses a standard ReLU function. The pooling is applied using max-pooling. The input image passes through a series of encoder blocks wherein, after each block, the number of channels scales up by a factor of 2. The final encoded activation map has 1024 channels. It passes through a series of decoder blocks wherein, after each block, the number of channels scales down by a factor of 2 shown in Fig. 2.

2.3 USegTransformer-P: Fully convolutional ensemble decoder

USegTransformer-P is a parallel stacked model of the transformer-based self-attention model and the UNet based model using a fully convolution-based decoder. The image is input into both the networks individually to output segmentation masks. These masks are stacked together and are convoluted using a 1 × 1 kernel. By using a 1 × 1 convolutional kernel, the model projects the best features by transfusing local and global features to produce a more accurate segmentation map. The flow of images in the network is mathematically represented as:

$$mg\in R^{H\times W\times 3}\Rightarrow feature\ matrix\in R^{H\times W\times 1} \tag{7}$$

$$img\in R^{H\times W\times 3}\Rightarrow feature\ matrix\in R^{H\times W\times 1} \tag{8}$$

$$R^{H\times W\times 1}+R^{H\times W\times 1}\Rightarrow concatenated\ matrix\in R^{H\times W\times 2} \tag{9}$$

$$R^{H\times W\times 2}\Rightarrow mask\in R^{H\times W\times 1} \tag{10}$$

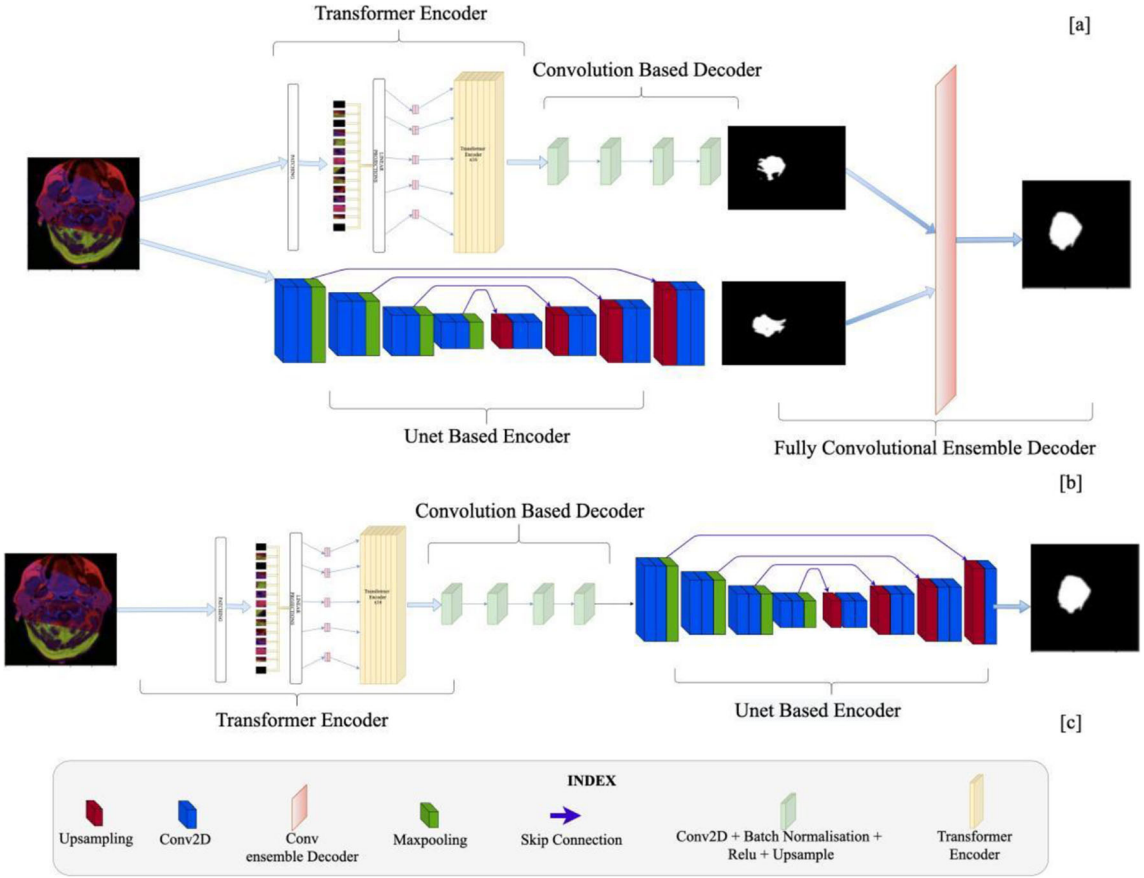


Fig. 2 Schematic of the proposed architecture (**a**) USegTransformer-P, (**b**) USegTransformer-S, and (**c**) index for the figure

where Eq. (7) represents the output of transformer-based encoder-decoder, (8) represents the outputs UNet inspired encoder decoder. The Eqs. (9) and (10) explain the image transformations in the fully convolution ensemble decoder. This model is trained using Algorithm-1.

Algorithm-1: Training of a USegTransformer-P Model

Input: A batch of image tensors
Output: Trained USegTransformer-P model

```
for step t = 1,2…T do
    for batch b = 1,2…N do
        x ← image tensor [b]
        x1 ← x
        for layer l = 1,2...L do
            x1 ← transformer_encoder(x1)
        End
        x1 ← convolution_based_decoder(x1)
        x2 ← unet_based_encoder_decoder_block(x)
        y_predicted[b] ← fully_conv_ensemble_decoder(x1,x2)
    End
    loss ← loss_function (y_predicted, y_true)
    Backpropagating loss error and weight update
End
```

The above training Algorithm-1 runs for T epochs, each step training on N batches making up the whole dataset. Each image in a batch runs through L layers of transformer encoder. The encoder applies attention to its input at each layer giving out a feature space that has considered and attended to the correlation between the image tokens. The final feature map is decoded using a convolutional decoder. In parallel, the image is also passed through a UNet based encoder-decoder to capture the spatial features and correlations. The two feature maps from both branches are combined and decoded to produce the output. The ensemble decoder combines the two feature spaces from the two branches and considers long range dependencies as well as spatially close dependencies to decide what features are more important where and then produces the segmented mask, which is an amalgamation of both branches. The output is compared to the ground truth and the error is backpropagated to update weights. The error is calculated using a loss function.

2.4 USegTransformer-S: Sequential stacking

The USegTransformer-S is a sequential stacked model of the UNet based model and inspired by the transformer-based self-attention model. The activation map produced by the transformer-based encoder-decoder is multichannel (in the proposed experiments, it has two channels). This multichannel

activation map is then input to the UNet based encoder-decoder. We intend to progressively down sample the number of channels in the output of each decoder. This helps in the transfusion of global features obtained by Transformer encoder and local feature obtained by Unet based encoder-decoder. The flow of the image in the network is mathematically represented as:

$$R^{\epsilon H\times W\times 3}\Rightarrow R^{\epsilon h\times w\times c}\Rightarrow R^{\epsilon H\times W\times 2}-1 \tag{11}$$

$$R^{\epsilon H\times W\times 2}\Rightarrow R^{\epsilon h'\times w'\times c'}\Rightarrow R^{\epsilon H\times W\times 1}-1 \tag{12}$$

The USegTransformer-S is trained using Algorithm-2.

Algorithm-2: Training of a USegTransformer-S Model

```
Input: A batch of image tensors
Output: Trained USegTransformer-S model
    for step t = 1,2…T do
        for batch b = 1,2…N do
            x ← image tensor [b]
            for layer l = 1,2...L do
                x ← transformer_encoder(x)
            end
            x ← convolution_based_decoder(x)
            x ← unet_based_encoder_decoder_block(x)
            y_predicted[b] ← x
        end
        loss ← loss_finction (y predicted, y_true)
        Backpropagating loss error and weight update
    end
```

The above training Algorithm-2 runs for T epochs, each step training on N batches making up the whole dataset. Each image in a batch runs through L layers of transformer encoder. The encoder applies attention to its input at each layer giving out a feature space that has considered and attended to the correlation between the image tokens. The final feature map is decoded using a convolutional decoder to give a multi-channel feature space. This feature space is further passed through a UNet based encoder decoder that capture the spatial features and relations in the high-level long-range feature space. The feature space also goes through a change in number of channels incorporating features at multiple levels. The output is, hence, a convoluted output of independent long-range features. This output is then compared to the ground truth and the error is backpropagated to update weights. The error is calculated using a loss function. Experimentally, as discussed in the next section, we have found that both ensembling methods perform better or at par with the previous state-of-the-art models wherein, the parallel model achieves superior results. Therefore, the fully convolutional ensemble decoder exploits the feature extraction abilities of both networks in a better way.

3 Experimental results

In order to investigate the capabilities of the proposed models, we test them on benchmark datasets. Each dataset chosen have a different set of challenges with a unique application in the field of medical imaging. The following section discusses in detail the various intricacies of each dataset as well as analyzes the results obtained on each dataset.

3.1 Dataset description

3.1.1 MRI dataset

The dataset [28] consists of MRI scans and corresponding segmentation masks of 110 patients obtained from the cancer imaging archive (TCIA). The patients are from the cancer Genome Atlas (TCGA) with low-grade glioma with fluid-attenuated inversion recovery (FLAIR) sequence. We have divided the chosen data into 70% training set(2838 images), 15% validation set(501 Images) and 15%(510 Images) testing set.

3.1.2 Lungs segmentation dataset

The dataset [29] is from the Lung Nodule Analysis (LUNA) competition from 2016. The competition aims to obtain automatic module detection algorithms. It contains 267 2D CT scans of lungs and their corresponding segmentation masks. We train the proposed models on 216 images (80%), validate the proposed models on 24 images (~10%), and test on 27 images (~10%).

3.1.3 Nuclei segmentation dataset

This dataset was part of the prestigious annual Kaggle data science bowl 2018 competition [30]. The dataset provides us with 670 images and along with them the segmented mask of each nucleus. The dataset contains images captured under varied conditions such as magnification, brightfield and fluorescence. The difference in quality images makes this dataset a true challenge for any deep learning model. In this paper, we use 603 (90%) images for training and 67 (10%) images for testing.

3.1.4 Skin lesion segmentation dataset

We use the dataset provided by the annual ISIC competition in 2018 [31, 32]. This dataset is meant to accomplish three tasks namely Lesion segmentation, lesion attribute detection and Disease classification. But since we intend to show our proposed models' proficiency in segmenting medical images, we concentrate on the first task. The dataset contains 2694 images and their corresponding masks. The masks are annotated by a

committee of experts. In this paper, we train our model on 75% of the data which is 2075 images and test the model's efficiency on 20% which is 518 images of the data and validated our model on 5% ~ 100 images.

3.1.5 COVID-19 CT scan segmentation dataset

This is a part of COVID-19 CT Image Segmentation competition hosted on Kaggle [33]. The dataset consists of 9 segmented axial volumetric CTs from Radiopedia. This dataset contains entire quantities and hence both positive and negative slices (373 out of the total of 829 slices have been evaluated by a radiologist as positive and segmented). These slices are then converted and normalized. Here the masks in the dataset consists of 4 classes or channels (ground glass, consolidations, lungs other, and background) out which 2 of those (ground glass and consolidations) have been used for evaluation of our proposed model. We trained the proposed model of 85% of these slices i.e., 704 slices and tested on 15% i.e., 125 slices.

The data sample splitting is conducted in the experiments for the different medical segmentation datasets as shown in Fig. 3. The split is made such that they are similar to splits in previous state-of-the-art and baseline results for the most appropriate comparison.

3.2 Pre-processing and training

All the experiments have been conducted in the PyTorch [34] framework on Google Collaboratory Pro. The GPU used was Tesla P-100. All the models were trained using the Ams grad [35] variant of Adam optimizer with the most appropriate learning rate and model hyperparameters according to the available resources. The data was pre-processed with a combination of some primary image augmentations like Random rotate, Random flip, Random Affine, and more. These augmentations were applied using Torchvision and Albumentations libraries in python. All the experiments were conducted under major resource constraints. Improvement in the following results might be seen by changing and tuning various variables given the required resources are available such as better GPUs, more disk space, and RAM.

3.2.1 Metrics

Three metrics are used for quantifying the performance of the proposed model which are Dice similarity coefficient, intersection over union and pixel accuracy [36]. The pixel accuracy is the percentage of pixels in the predicted mask that match the expected pixel class in the ground truth mask. It is the most simple and primitive metric which is prone to class imbalance. The mathematical formula to determine pixel accuracy is presented as:

$$Pixel\ Accuracy = \frac{N_{TP} + N_{TN}}{N_{TP} + N_{TN} + N_{FP} + N_{FN}} \tag{13}$$

where, N_{TP}, N_{TN}, N_{FP}, and N_{FN} are the correctly classified pixels as Class A, correctly classified pixels as not Class A, incorrectly classified pixels as class A, and incorrectly classified pixels as not Class A, respectively. The intersection over union (IoU) is one of the widely accepted metrics of measuring the efficacy of the segmentation algorithm. It measures the overlap between the ground truth and the predicted masks by dividing the area of overlap by the area of union. The mathematical expressions to determining IoU is given as:

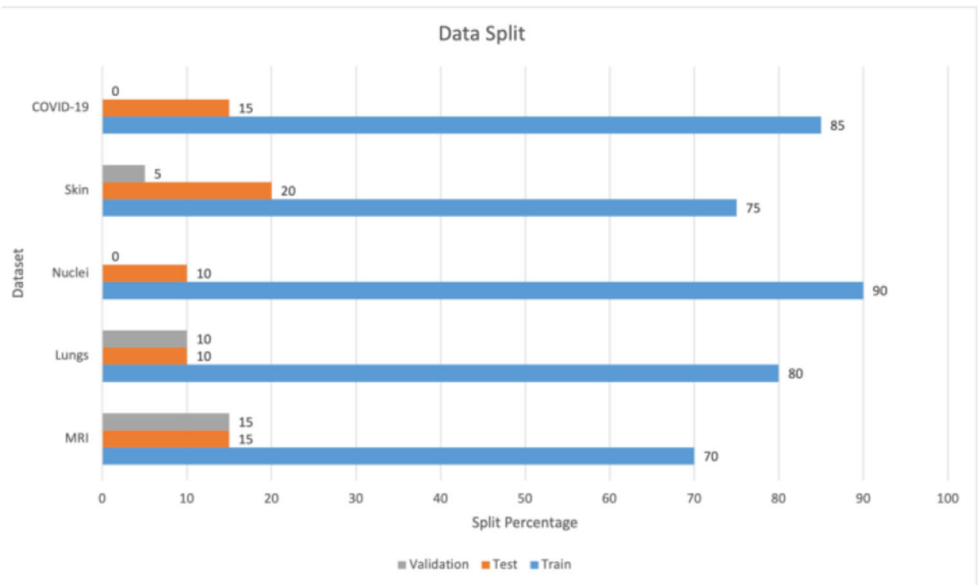


Fig. 3 Split in datasets

$$IoU = \frac{|N_A \cap N_B|}{|N_A \cup N_B|} \tag{14}$$

where, N_A, N_B, $N_A \cap N_B$, and $N_A \cup N_B$ are the pixels of Image A, pixels of Image B, area of overlap and area of union, respectively. The Dice coefficient is one of the best metrics to measure the efficacy of segmentation algorithm. It measures the similarity between the ground truth and predicted masks by dividing the number of overlapping pixels by total number of pixels in both images and multiplying the results by two. Equation (15) shows the mathematical expressions for determining Dice.

$$Dice = \frac{2*|N_A \cap N_B|}{|N_A| + |N_B|} \tag{15}$$

where, N_A, N_B, and $N_A \cap N_B$ are the pixels of image A, pixels of image B, and the area of overlap, respectively.

3.2.2 Loss functions

Binary cross entropy loss function is a standard loss function used in classification and segmentation task. The primary advantage of binary cross entropy loss function is that it provides smooth loss curves which contributes towards faster training of models.

$$BCE = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot \log\left(\widehat{y}_i\right) + (1-y_i) \cdot \log\left(1-\widehat{y}_i\right) \tag{16}$$

The Dice Loss is a loss function that makes the model strive towards producing images which are similar to the ground truth. It is slowly becoming a popular choice owing to its inherent maximization of dice coefficient and its salubrious effect on class imbalance. Mathematically dice loss is calculated by simply subtracting dice coefficient from 1.

$$Dice\ Loss = 1 - \sum_{i=0}^{i=N} \frac{2*\left|y_i \cap \widehat{y}_i\right|}{|y_i| + \left|\widehat{y}_i\right|} \tag{17}$$

All the models are trained using the standard binary cross-entropy (BCE) Dice loss, since both BCE loss and Dice loss when combined are observed to have a symbiotic relationship. The combined loss has relatively smooth loss curves owing to the component contributed by BCE loss as well Dice Loss's inherent tendency to maximize Dice coefficient and handle class imbalance which is prevalent in medical imaging datasets.

$$Combined\ Loss = -\frac{1}{N}\sum_{i=0}^{N} y_i \cdot log\left(\widehat{y}_i\right) + (1-y_i) \cdot log\left(1-\widehat{y}_i\right) + 1 - \sum_{i=0}^{i=N} \frac{2*\left|y_i \cap \widehat{y}_i\right|}{|y_i| + \left|\widehat{y}_i\right|} \tag{18}$$

3.3 Results

In this subsection, we have discussed in detail as well as analyze the prediction made by the models on various datasets. The quantitative analysis is done on the aforementioned metrics however, only one metric could be used for comparative analysis with prior models since most datasets were a part of a competition which required reporting that specific metric. Further visual predictions made by proposed model are also presented in order to analyze the performance of the proposed model qualitatively.

3.3.1 Brain tumor segmentation

One of the most crucial application of medical image segmentation is Brain tumor segmentation from Brain MRI. We choose Brain MRI LGG dataset from Kaggle to evaluate our proposed model. From Table 1, it is depicted that the USegTransformer-P performs the best across all segmentation metrics, that is, Dice score, Intersection over union and accuracy. The USegTransformer-P achieves an accuracy of 99.71 while achieve an IOU score of 0.9734 and a dice of 0.8934 while converging approximately at the same time with transformer backbone as seen in Fig. 4b. Figure 4c shows the validation curve (BCE Dice Loss Vs Epoch) for all three models, here even though the transformer backbone outperforms USegTransformer-S and USegTransformer-P on BCE Dice loss, yet USegTransformer-S and USegTransformer-P achieve a better dice coefficient and a better IoU score on validation split as well. Furthermore, these results can be visually analyzed from Fig. 4a. The transformer Backbone also performs efficiently achieving a dice score 8770, an IOU of 0.9381 and an accuracy of 0.9961. It is interesting to note that

Table 1 Key performance indicator of MRI segmentation

Model	Dice	IoU	Accuracy
U-Net [17]	0.8200	–	–
SynthSeg [37]	0.8610	–	–
Transformer Backbone	0.8770	0.9381	0.9961
USegTransformer-P	**0.8934**	**0.9746**	**0.9971**
USegTransformer-S	**0.8504**	**0.9634**	**0.9954**

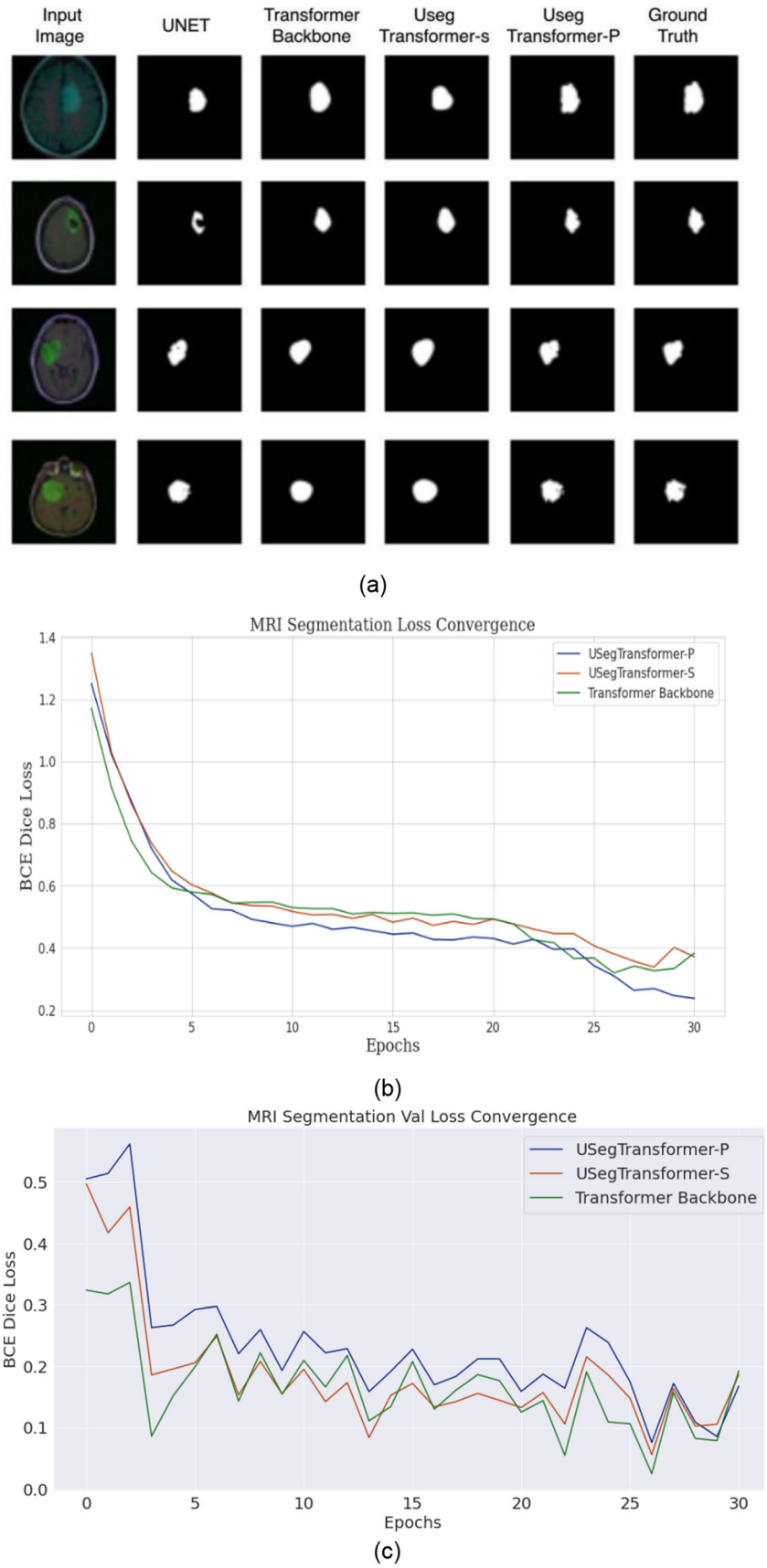


Fig. 4 The Brain MRI segmentation (**a**) visual depiction, (**b**) train loss convergence and (**c**) validation Loss

on this particular dataset transformer backbone performs better than the USegTransformer-S which is a deeper model. The USegTransformer-S achieves a dice score of 0.8504, an IOU of 0.9634 and an accuracy of 0.9954.The subpar performance of the USegTransformer-S on this dataset can be attributed to the lack of images in the dataset in comparison to the depth of the model. As shown by Table 1, the model outperforms state-of-the-art of models by a considerable overhead.

3.3.2 Lung nodule segmentation

Another crucial application where an artificial intelligence can revolutionize medical diagnosis is Lung Nodule analysis. We choose the dataset provided in the famous challenge LUNA to evaluate the proposed model's efficacy in segmentation of lung nodules. As presented in Table 2, the USegTransformer-P performs the best amongst Transformer Backbone, the USegTransformer-S and USegTransformer-P. The USegTransformer-P achieves a Dice score of 0.9807, an IOU of 0.9462 and an accuracy of 0.9913. While the USegTransformer-S achieves a Dice score 0.9777, an IOU of 0.9316 and an accuracy of 0.9913. Figure 4b and Fig. 4c shows the convergence of the models on BCE Dice Loss. The loss convergence and the results on the test split clearly show how USegTransformer-P generalizes on the dataset. The results are visualized in Fig. 5a. The Transformer Backbone, which is the least complex model achieves a Dice score of 0.9582, an IOU of 0.8946 and an accuracy of 0.9816. It can be observed from the Table 2 that the proposed model either outperforms most of the state-of-the-art models while performing at par with others.

Additionally, we performed 3-fold cross validation on the lungs segmentation dataset resulting in the model to be trained and tested on the data being divided into 2:1 ratio (66% training and 33% testing split). The results in Table 3 showcases that the model upholds its high dice, IoU and accuracy metrics with very low standard deviation across the 3 folds. This proves that the proposed architectures are robust to the data split.

Table 2 Key performance parameters of lungs segmentation

Model	Dice	IoU	Accuracy
U-Net [17]	–	–	0.9872
RU-Net [38]	–	–	0.9836
CE-Net [24]	–	–	0.9900
ET-Net [39]	–	–	0.9868
BCDU-Net [21]	–	–	0.9946
Transformer Backbone	0.9582	0.8946	0.9816
USegTransformer-P	**0.9807**	**0.9462**	**0.9913**
USegTransformer-S	**0.9777**	**0.9316**	**0.9894**

3.3.3 Skin lesion segmentation

The Skin Lesion Segmentation is a vital process in medical diagnosis since it forms the basis for more complex analysis. Table 4 presents the key performance parameters of proposed model on the ISIC 2018 competition dataset. The USegTransformer-P is the most effective in skin lesion segmentation achieving an accuracy of 0.9514, an IOU of 0.8672 and a Dice score of 0.8701. Figure 6a provides a visual depiction of these results. Further, the USegTransformer-P shows better and faster convergence and does not overfit on the data as evident Fig. 6b and Fig. 6c. The performance of Transformer Backbone and USegTransformer-S is similar. The USegTransformer-S achieves a Dice score of 0.8420, an IOU of 0.8374 and an accuracy of 0.9431. While the Transformer Backbone achieves a Dice score of 0.8503, an IOU of 0.8511 and an accuracy of 0.9447. As evident by Table 3, the best model proposed in this paper is able to significantly outperform various state of segmentation architectures.

3.3.4 Nuclei segmentation

An automation in the field of nuclei segmentation can prove to be a game changer in bio-medical research and thus is an important area of research. We evaluate the efficacy of proposed model in segmenting nuclei using the dataset provided in Kaggle 2018 data science bowl. From Table 5, it is evident that the proposed model that is USegTransformer-P has outperformed all the other models by achieving a Dice score of 0.9004, an accuracy of 0.9761 and IoU of 0.8470. These exemplary results can be seen in Fig. 7a. On the other hand, USegTransformer-S has achieved a dice score of 0.8517, accuracy of 0.9653 and an IoU of 0.7949. The Transformer Backbone has however outperformed the USegTransformer-S model, achieved a Dice score of 0.8780, accuracy of 0.9694 and IoU of 0.8287. The results explained above are reiterated by analyzing loss curves in Fig. 7b which shows that the USegTransformer-P is able to converge better.

Further, we performed 3-fold cross validation on the nuclei segmentation dataset resulting in the model to be trained and tested on the data being divided into 2:1 ratio (66% training and 33% testing split). The results in Table 6 showcase that the model upholds its high dice, IoU and accuracy metrics with very low standard deviation across the 3 folds. This proves that the proposed architectures showcase high degree of robustness.

3.3.5 Covid −19 CT-scan segmentation

An automatic segmentation also proves to be a boon to the bane of humanity that is Covid-19 pandemic. A major effect of Covid −19 is consolidation in chest CT. We believe that

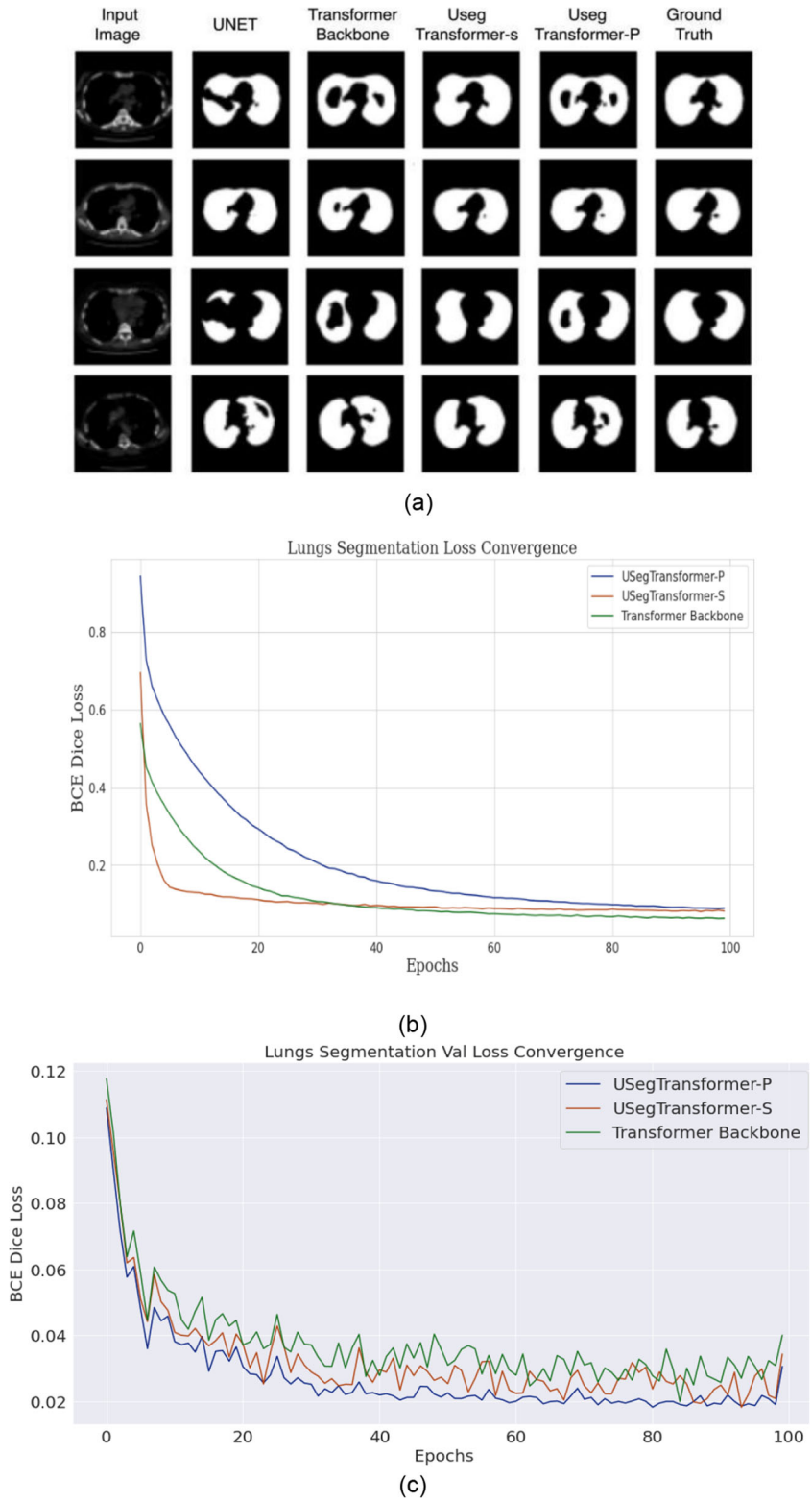


Fig. 5 The lungs segmentation (**a**) visual depiction (**b**) train loss convergence and (**c**) validation loss convergence

segmenting these consolidations will lead to better qualitative analysis of patients. Table 7 shows the quantitative analysis of the COVID-19 Consolidation mask dataset. The USegTransformer-P achieved a Dice score of 0.8295, an accuracy of 0.9981 and an IoU of 0.9685. The USegTransformer-P has achieved a Dice score of 0.6811, accuracy of 0.9919 and IoU of 0.9218 for ground class predictions. Figure 8a provides a qualitative analysis of these results.

Table 3 3-Fold Cross Validation Performance on lungs segmentation

Model	Mean Metric ± Std Deviation		
	Dice	IoU	Accuracy
USegTransformer-P	**0.9728±0.0052**	**0.9333±0.0046**	**0.9901±0.0013**
USegTransformer-S	**0.9601±0.0083**	**0.9256±0.0066**	**0.9811±0.0023**

The trend is also reiterated in the loss convergence graph in Fig. 8b.

Moreover, we performed 3-fold cross validation on the COVID-19 segmentation dataset resulting in the model to be trained and tested on the data being divided into 2:1 ratio (66% training and 33% testing split). The results in Table 8 showcase that the model upholds its high dice, IoU and accuracy metrics with very low standard deviation across the 3 folds. This proves that the proposed architecture is robust to the data split.

4 Discussions

The medical image segmentation is one of most crucial tasks in the diagnosis obtained from analyzing medical images. The current trend in automated image segmentation is to improve the performance of traditionally used fully convolution networks by optimizing skip connection, increasing the strength of convolution encoders. However, one area where very limited research has been done is making deep learning models account long range dependencies. We believe that such consideration of such long-range dependencies holds the key to a fully independent deep learning based medical segmentation system. In the proposed work, we try to account long range dependencies by virtue of transformer encoders. The two proposed models in this study utilize both, spatial and global features and transfuse them in two unique manners. The first model, the USegTransformer-P theoretically lets the image run through the transformer as well as the UNet encoder

Table 4 Performance on skin lesion segmentation

Model	Dice	IoU	Accuracy
BCDU-Net [21]	–	–	0.937
U-Net [17]	–	–	0.890
R2U-Net [38]	–	–	0.880
Attention R2U-Net [38]	–	–	0.904
Attention U-Net [40]	–	–	0.897
Transformer Backbone	0.8503	0.8511	0.9447
USegTransformer-P	**0.8701**	**0.8671**	**0.9514**
USegTransformer-S	**0.8420**	**0.8374**	**0.9431**

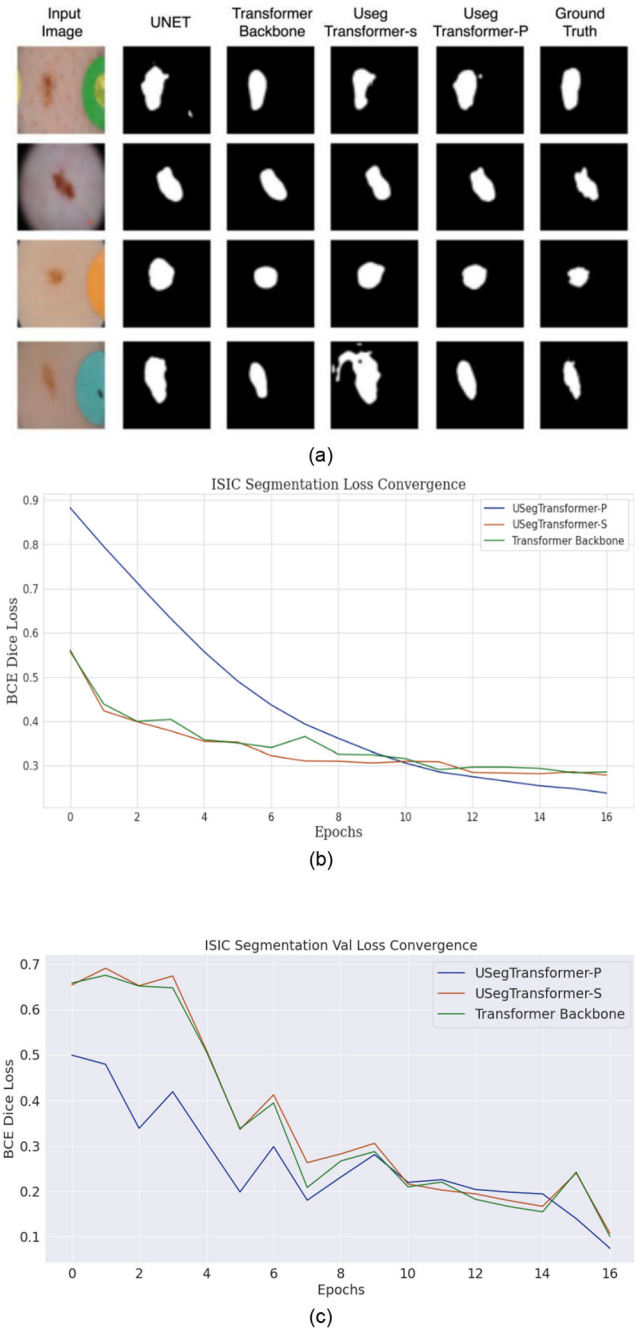


Fig. 6 The skin lesion segmentation (**a**) visual depiction, (**b**) train loss convergence and (**c**) validation loss convergence

decoders and finally combines and chooses between the local and global features through the novel ensemble decoder, from their respective feature maps. On the other hand, the second model, USegTransformer-S, first outputs a multichannel feature map considering long range features which is then used to find local features. This leads to an affective transfusion of both types of features. On analysing Fig. 9, specifically the third row, the effects that global features induce in the quality of masks becomes clearer. We can observe from Fig. 9, that the masks produced in UNET have very detailed boundaries

Table 5 Key performance indicators on nuclei segmentation

Model	Dice	IoU	Accuracy
U-Net [17]	0.7573	–	–
U-Net++ [20]	0.8974	–	–
32xU-Net/FPN (top coders) [30]*	–	–	0.6316
1xFC-FPN (Jacobkie) [30]*	–	–	0.6147
1xMask-RCNN (Deep Retina) [30]*	–	–	0.6140
Transformer Backbone	0.8780	0.8287	0.9694
USegTransformer-P	**0.9004**	**0.8470**	**0.9761**
USegTransformer-S	**0.8517**	**0.7949**	**0.9653**

*Winning teams from the original competition

while the masks produced by transformer backbone model have plain boundary. However, masks produced UNET are not able to emulate the broader shape of mask as effectively as the transformer. Such observation can be attributed to the type of features each model is using, UNET uses convolution which is able to exploit the local information and explicitly determine sharp edges present in boundaries while transformer which use attention mechanism to form correlation between

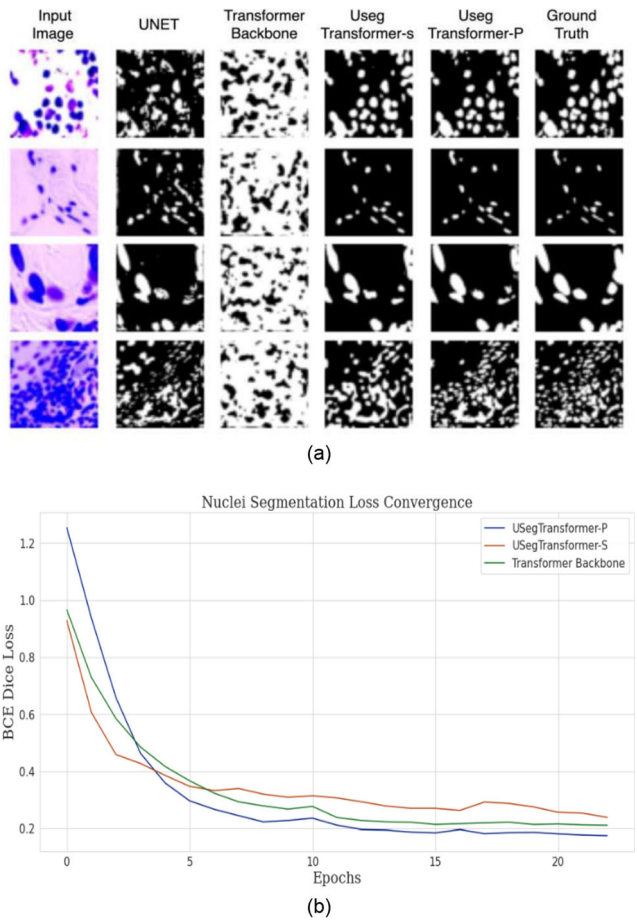


Fig. 7 The nuclei segmentation (**a**) visual depiction, and (**b**) loss convergence

Table 6 3-Fold Cross Validation Performance on nuclei segmentation

Model	Mean Metric ± Std Deviation		
	Dice	IoU	Accuracy
USegTransformer-P	**0.8887±0.0095**	**0.8351±0.0073**	**0.9703±0.0031**
USegTransformer-S	**0.8311±0.0114**	**0.7790±0.0098**	**0.9588±0.0054**

input has a large scope which leads to better understanding of general shape. The mask produced by USegTransformer-P.

An important trend, we try to set through the proposed model is to expand the use of deep learning models to segmentation for detection of COVID-19 which was earlier on limited to classification problem statements. It has been illustrated that the USegTransformer-S model which is a product of sequential stacking performed at par with the existing models and didn't perform better than that of the USegTransformer-P model. Since, the sequential model is a relatively deeper model and for the benchmark datasets, we have smaller amounts of data. Hence, we hypothesize that this model holds the capacity to work much better if it is trained on huge industrial datasets with resources much greater than ours. In the proposed research work, we have been able to propose a novel model that holds various advantages over other networks that makes the proposed models much more suitable for the task of medical image segmentation. As explained earlier, the proposed models have the capacity to outperform other previous state-of-the-art networks qualitatively and quantitatively. It has the capacity to produce segmentation masks with accurate high-level form and precise boundaries and spacings which entitles these predicts to be more trustworthy and reliable for real world applications. On the flip side, it is important to note that the USegTransformer-P and USegTransformer-S are large computationally complex models that require huge training data to function at their best capacities.

5 Conclusion

In this paper, we have proposed an end-to-end deep learning frameworks for segmenting medical images named USegTransformer-P and USegTransformer-S which are capable of aiding medical professionals as well as accelerating the

Table 7 Key Performance indicators on COVID-19 segmentation

Model	Mask Feature	Dice	IoU	Accuracy
USegTransformer-P	Consolidations	0.8295	0.9685	0.9981
USegTransformer-P	Ground Glass	0.6811	0.9218	0.9919

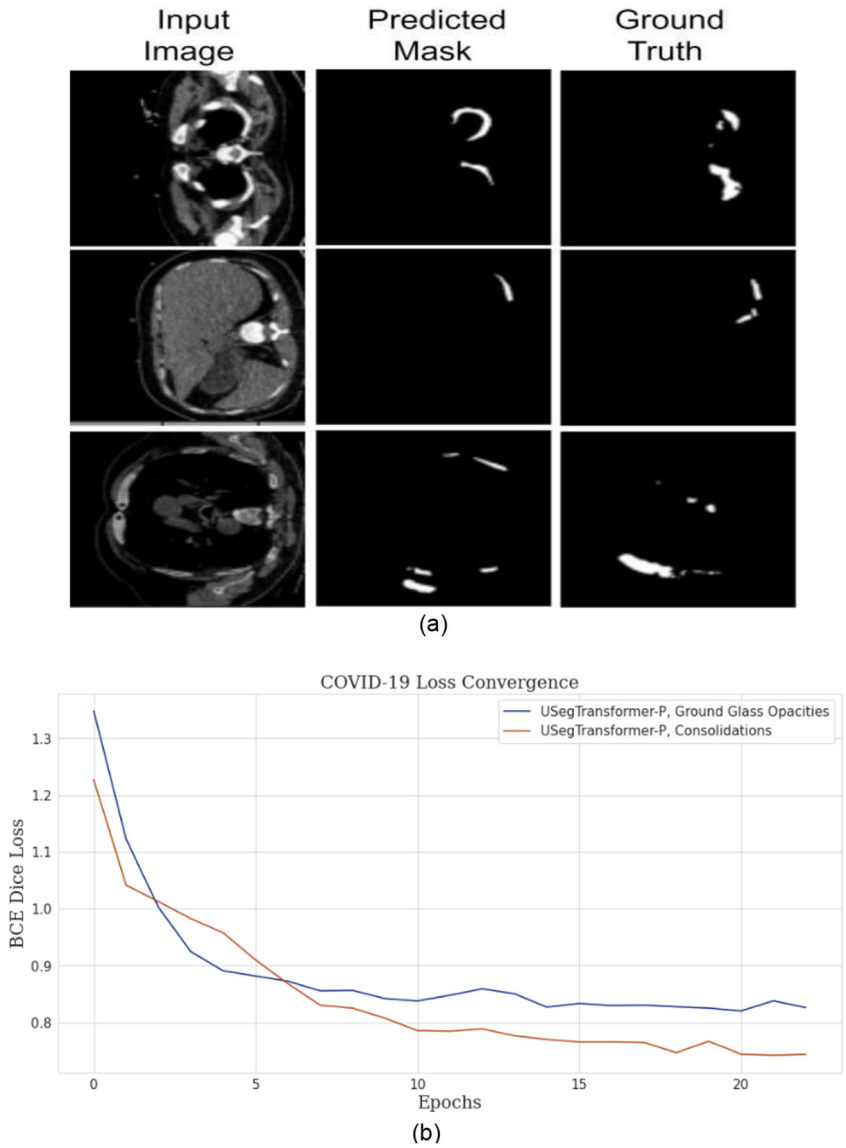


Fig. 8 The COVID-19 CT scan segmentation (**a**) visual depiction, and (**b**) loss convergence

process of medical diagnoses. Further, we illustrated the efficiency of utilizing transformer-based encoding and FCN based encoding together in the model. Also, we have presented two different techniques of stacking FCN-based segmentation models and transformer-based segmentation models. Furthermore, we have presented the efficiency of proposed model by evaluating it on varied benchmark datasets like LGG, LUNA, ISIC, and Data Science Bowl 2018 Dataset where the proposed model USegTransformer-P beat the current state of the models by achieving accuracies of 99.71%, 99.13%, 95.14% and 97.61% as well as USegTransformer-S achieved accuracies of 99.54%, 98.94%, 94.31% and 96.53%, respectively. Moreover, we have proved that the localized features obtained from FCN based networks and global context features obtained from transformer-based networks complement each other in improving a model's segmentation ability of medical datasets. By the virtue of these qualitative and quantitative

Table 8 3-Fold Cross Validation Performance on COVID-19 segmentation

Mask Feature	Mean Metric ± Std Deviation		
	Dice	IoU	Accuracy
USegTransformer-P (Consolidations)	**0.7854±0.0141**	**0.9746±0.0029**	**0.9989±0.0001**
USegTransformer-P (Ground Glass)	**0.6776±0.0159**	**0.9149±0.0131**	**0.9910±0.0014**

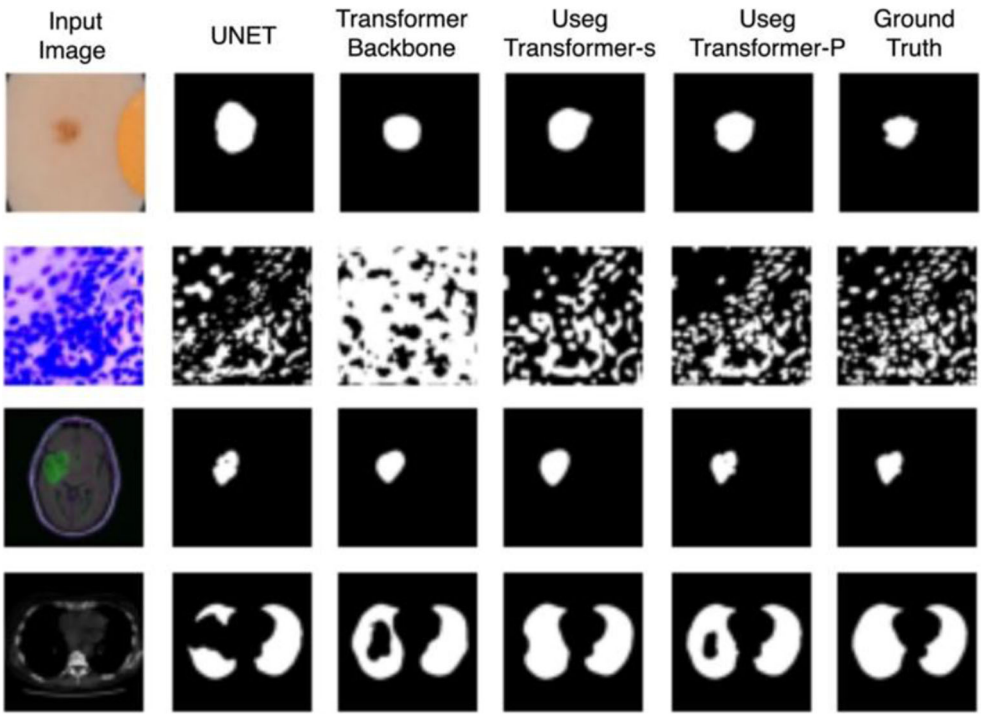


Fig. 9 Effects of global and local features

improvements, the proposed models are trustable and appropriate for real-world clinical applications. These predictions have the capacity to work in a diagnosis system to help analyze medical scans and reports to facilitate medical workers in the process of making health infrastructure accessible, available, and fruitful potential users.

In the future, experiments can be conducted on developing a more complex patching strategy in order to make the process of patching dynamic. Currently, the patching in a transformer branch is rigid and depends on manually parameterized sizes which can hold the transformer back in however more optimal patches can produce better results. In order to make the process of patching dynamic, we have to develop and conduct a more complex patching strategy. Further, an experiment could be conducted to develop an attention which is optimized for vision tasks instead of the self-attention mechanism which is a direct translation of the concept originally establish in the NLP domain that forces the transformer to look at a series of image tokens as a sentence with positional encoding. Moreover, different variants of UNet could be used to experiment with different types of feature extraction since the improvement of image segmentation is by respecting different types of features. Further, the full potential of USegTransformer-S can be analyzed by evaluating the model on more complex and large datasets.

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

References

1. Ganguly D, Chakraborty S, Balitanas M, Kim TH (2010) Medical imaging: a review. Commun Comput Inf Sci 78(CCIS):504–516. https://doi.org/10.1007/978-3-642-16444-6_63
2. Cao H, Liu H, Song E, Hung CC, Ma G, Xu X, Jin R, Lu J (2020) Dual-branch residual network for lung nodule segmentation. Appl Soft Comput J 86:105934. https://doi.org/10.1016/j.asoc.2019.105934
3. Hashemzehi R, Mahdavi SJS, Kheirabadi M, Kamel SR (2020) Detection of brain tumors from MRI images base on deep learning using hybrid model CNN and NADE. Biocybern Biomed Eng 40:1225–1232
4. Garcia-Arroyo JL, Garcia-Zapirain B (2019) Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding. Comput Methods Prog Biomed 168:11–19. https://doi.org/10.1016/j.cmpb.2018.11.001
5. Nogueira-Rodríguez A, Domínguez-Carbajales R, López-Fernández H, Iglesias Á, Cubiella J, Fdez-Riverola F, Reboiro-Jato M, Glez-Peña D (2021) Deep neural networks approaches for detecting and classifying colorectal polyps. Neurocomputing 423:721–734. https://doi.org/10.1016/j.neucom.2020.02.123
6. Chlebus G, Schenk A, Moltz JH, van Ginneken B, Hahn HK, Meine H (2018) Automatic liver tumor segmentation in CT with fully convolutional neural networks and object-based postprocessing. Sci Rep 8:15497
7. Lal S, Das D, Alabhya K, Kanfade A, Kumar A, Kini J (2021) NucleiSegNet: robust deep learning architecture for the nuclei segmentation of liver cancer histopathology images. Comput Biol Med 128:104075. https://doi.org/10.1016/j.compbiomed.2020.104075
8. Sharma N, Aggarwal LM (2010) Automated medical image segmentation techniques. J Med Phys 35(1):3–14. https://doi.org/10.4103/0971-6203.58777
9. Ramesh N, Yoo JH, Sethi IK (1995) Thresholding based on histogram approximation. IEE Proc Vision, Image Signal Process 142:271. https://doi.org/10.1049/ip-vis:19952007

10. Sharma N, Ray A, Sharma S, Shukla KK, Pradhan S, Aggarwal LM (2008) Segmentation and classification of medical images using texture-primitive features: application of BAM-type artificial neural network. J Med Phys 33:119–126. https://doi.org/10.4103/0971-6203.42763
11. Gletsos M, Mougiakakou SG, Matsopoulos GK, et al (2003) A computer-aided diagnostic system to characterize CT focal liver Lesions: Design and Optimization of a Neural Network Classifier. IEEE Trans Inf Technol Biomed. https://doi.org/10.1109/TITB.2003.813793
12. Zheng X, Lei Q, Yao R, Gong Y, Yin Q (2018) Image segmentation based on adaptive K-means algorithm. Eurasip J Image Video Process 2018. https://doi.org/10.1186/s13640-018-0309-3
13. Ahmadi N (2020) A hybrid intelligent approach for image segmentation and feature extraction using fuzzy clustering, lattice boltzmann and GLDM techniques. J Soft Comput Decis Support Syst 7:1–5. https://doi.org/10.5815/ijigsp.2012.06.01
14. Bezdek JC, Ehrlich R, Full W (1984) FCM: the fuzzy c-means clustering algorithm. Comput Geosci 10:191–203. https://doi.org/10.1016/0098-3004(84)90020-7
15. Krizhevsky A, Sutskever I, Hinton GE (2017) ImageNet classification with deep convolutional neural networks. Commun ACM 60:84–90. https://doi.org/10.1145/3065386
16. Ciresan DC, Giusti A, Gambardella LM, Schmidhuber J (2012) Deep neural networks segment neuronal membranes in electron microscopy images. In: NIPS, pp 2852–2860
17. Ronneberger O, Fischer P, Brox T (2015) U-Net: convolutional networks for biomedical image segmentation. MICCAI. https://doi.org/10.48550/arXiv.1505.04597
18. Ibtehaz N, Rahman MS (2020) MultiResUNet: rethinking the U-net architecture for multimodal biomedical image segmentation. Neural Netw 121:74–87. https://doi.org/10.1016/j.neunet.2019.08.025
19. Lou A, Guan S, Loew M (2020) DC-UNet: rethinking the U-net architecture with dual channel efficient CNN for medical images segmentation. arXiv. https://doi.org/10.48550/arXiv.2006.00414
20. Zhou Z, Siddiquee MMR, Tajbakhsh N, Liang J (2020) UNet++: redesigning skip connections to exploit multiscale features in image segmentation. IEEE Trans Med Imaging 39:1856–1867. https://doi.org/10.1109/TMI.2019.2959609
21. Azad R, Asadi-Aghbolaghi M, Fathy M, Escalera S (2019) Bi-directional ConvLSTM U-net with densley connected convolutions. In: Proceedings - 2019 international conference on computer vision workshop, ICCVW 2019. https://doi.org/10.48550/arXiv.1909.00166
22. Jha D, Riegler MA, Johansen D et al (2020) DoubleU-net: a deep convolutional neural network for medical image segmentation. In: Proceedings - IEEE Symposium on Computer-Based Medical Systems. https://doi.org/10.48550/arXiv.2006.04868
23. Chen L, Bentley P, Mori K, Misawa K, Fujiwara M, Rueckert D (2018) DRINet for medical image segmentation. IEEE Trans Med Imaging 37:2453–2462. https://doi.org/10.1109/TMI.2018.2835303
24. Gu Z, Cheng J, Fu H, Zhou K, Hao H, Zhao Y, Zhang T, Gao S, Liu J (2019) CE-net: context encoder network for 2D medical image segmentation. IEEE Trans Med Imaging 38:2281–2292. https://doi.org/10.1109/TMI.2019.2903562
25. Dosovitskiy A, Beyer L, Kolesnikov A et al (2020) An image is worth 16x16 words: transformers for image recognition at scale. ICLR 2021. https://doi.org/10.48550/arXiv.2010.11929
26. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser Ł, Polosukhin I (2017). Attention is all you need. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NIPS'17). Curran Associates Inc., Red Hook, NY, USA, pp 6000–6010
27. Nishio M, Nagashima C, Hirabayashi S, Ohnishi A, Sasaki K, Sagawa T, Hamada M, Yamashita T (2017) Convolutional auto-encoders for image denoising of ultra-low-dose CT. Heliyon 3:e00393. https://doi.org/10.1016/j.heliyon.2017.e00393
28. Brain MRI segmentation | Kaggle (n.d.) https://www.kaggle.com/mateuszbuda/lgg-mri-segmentation
29. Finding and Measuring Lungs in CT Data | Kaggle (n.d.) https://www.kaggle.com/kmader/finding-lungs-in-ct-data/data/
30. Caicedo JC, Goodman A, Karhohs KW, Cimini BA, Ackerman J, Haghighi M, Heng CK, Becker T, Doan M, McQuin C, Rohban M, Singh S, Carpenter AE (2019) Nucleus segmentation across imaging experiments: the 2018 data science bowl. Nat Methods 16:1247–1253. https://doi.org/10.1038/s41592-019-0612-7
31. Codella NC, Rotemberg VM, Tschandl P, Celebi ME, Dusza SW, Gutman D, Helba B, Kalloo A, Liopyris K, Marchetti MA, Kittler H, Halpern AC (2019) Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC). ArXiv, abs/1902.03368
32. Tschandl P, Rosendahl C, Kittler H (2018) Data descriptor: the HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Sci Data 5:180161. https://doi.org/10.1038/sdata.2018.161
33. COVID-19 - Medical segmentation (n.d.) http://medicalsegmentation.com/covid19/
34. Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N, Antiga L, Desmaison S, Köpf A, Yang E, DeVito Z, Raison M, Tejani A, Chilamkurthy S, Steiner B, Fang L, Bai J, Chintala S (2019) PyTorch: An imperative style, high-performance deep learning library. NeurIPS. https://doi.org/10.48550/arXiv.1912.01703
35. Reddi SJ, Kale S, Kumar S (2018) On the convergence of Adam and beyond. In: 6th international conference on learning representations, ICLR 2018 - conference track proceedings. https://doi.org/10.48550/arXiv.1904.09237
36. Rizwan I, Haque I, Neubert J (2020) Deep learning approaches to biomedical image segmentation. Informatics Med Unlocked 18:100297/1–100297/10029711. https://doi.org/10.1016/j.imu.2020.100297
37. Billot B, Greve DN, van Leemput K et al (2020) A learning strategy for contrast-agnostic MRI segmentation.Proceedings of the Third Conference on Medical Imaging with Deep Learning 121:75-93. https://doi.org/10.48550/arXiv.2003.01995
38. Alom MZ, Yakopcic C, Hasan M, Taha TM, Asari VK (2019) Recurrent residual U-net for medical image segmentation. J Med Imaging 6:1. https://doi.org/10.1117/1.jmi.6.1.014006
39. Zhang Z, Fu H, Dai H, Shen J, Pang Y, Shao L (2019) ET-Net: A generic edge-aTtention guidance network for medical image segmentation. In: Medical Image Computing and Computer Assisted Intervention – MICCAI 2019. Springer-Verlag, Berlin, Heidelberg, pp 442–450. https://doi.org/10.1007/978-3-030-32239-7_49
40. Oktay O, Schlemper J, Folgoc LL, Lee MJ, Heinrich MP, Misawa K, Mori K, McDonagh SG, Hammerla NY, Kainz B, Glocker B, Rueckert D (2018) Attention U-Net: learning where to look for the pancreas. ArXiv, abs/1804.03999

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

International Journal of Multidisciplinary Research and Growth Evaluation.

Solar Energy: A source for E-vehicles

Kaseri Abdul Rehman [1*], Naveen Kumar [2], JP Kesari [3]
[1-3] Department of Mechanical Engineering, Delhi Technological University, Bawana Road, Delhi, India

* Corresponding Author: **Kaseri Abdul Rehman**

Article Info

ISSN (online): 2582-7138
Volume: 03
Issue: 02
March-April 2022
Received: 22-03-2022;
Accepted: 09-04-2022
Page No: 518-522

Abstract
Demand and availability of power are growing day by day throughout the world. Renewable energy sources will play an important role in the future as normal energy sources such as coal and petroleum are limited and depleted. In this case, the power problem is a very important problem and therefore needs to be solved. Basically, it is to ensure the availability to increase generation or install additional generators. Otherwise, a demand side management (DSM) approach must be adopted. In the current situation, fuel consumption for domestic transport is also showing an increasing threat in the near future. The negative effects of the practice of such vehicles as well means that it leads to environmental pollution and health issues related to habitats. In line with this future transport system E-cars are always worried about environmental pollution. The drawback of this method of transporting energy consumption by moving it will also create an additional burden on the setting of the existing power system and infrastructure. But power consumption and controls in electric vehicles (EVs) and the latest technology promotes and supports green transport systems. The E-Vehicle Charger plays a role in reducing dependence on mineral oil and is the natural emergence of our most powerful infrastructure. Consumer incentives as well it is important to make the purchase of an electric car and a quick local charging station affordable and attractive. To address this issue many developed countries have already taken action to remove solar photovoltaic conversion systems for storage and transfer to the grid. The current rate of use of E vehicles is 70% compared to a decade ago. The concept of solar energy is basically, released as heat or light energy. But the storage system is the bottleneck issue in this case due to technology, economic barriers as well as challenges. Power electronics connect the male E-car battery to be charged to the grid if you do not have a solar power source known as a car grid (G2V). Another mode says to transfer stored power from storage battery to grid (V2G) as and when the grid receives lack of ability to meet its responsibilities. Then solar PV systems are installed on the roof of the car to generate electricity and are used for its operation with battery support.
Therefore, E cars equipped with solar PV panels can transmit even electricity to the grid (V2G). Therefore, proper design and application can prove solar photovoltaic source as a promising source for this E car car app in the near future. This creates an opportunity for engineers and businesses.

Keywords: Solar Energy, E-vehicles, Demand

Introduction
The main objectives of the concept “Solar energy a promising source for E-vehicles” is to explore the ability and to utilize solar energy for E-vehicles thus reducing the fossil fuel consumption and protecting the environment. The conventional vehicles and potential for the use of alternative technologies in automobiles such as electric/hybrid vehicles need to be worked out or develop a plug-in hybrid electric vehicle by converting available conventional vehicles with a suitable motor and battery. It is also necessary to know why E-vehicle couldn’t get enough consumer attraction. Need for E-vehicles marketing and production can be done and problems can be solved by applying suitable engineering solution techniques considering consumer requirements and cost constraints. Consumers should be aware of the CO2 emissions caused by fossil fuel based vehicles and the initiatives taken by the government for promoting E-vehicles by giving subsidies.

The demand for electricity is growing exponentially, which is very important in all aspects of modern life. This is evidenced by the exponential growth of the world's population. Therefore, energy sources will play an important role. In this regard, current practices have the potential and priorities to reduce greenhouse gas emissions and improve residential, office and industrial energy efficiency, energy marketing, energy management, energy conservation and energy security. They are also working to find viable and affordable alternative energy sources by developing cleaner and more efficient vehicles and systems, energy policies and energy strategies. Over Extraction or depletion of fossil fuels and environmental degradation are major problems facing the modern world? In order to overcome these problems, interest in renewable energy sources has recently increased due to environmental benefits. The competition between fossil fuels and renewable fuels is important for the foreseeable future.

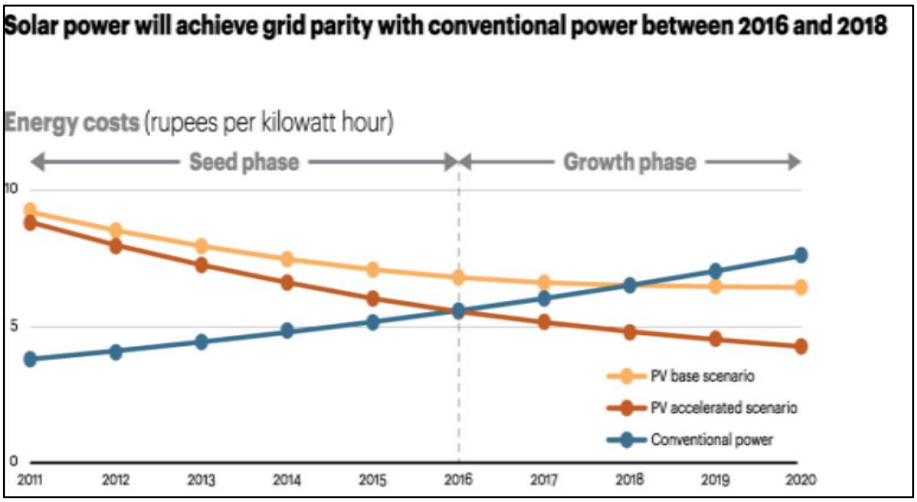


Fig 1: Variation of energy costs

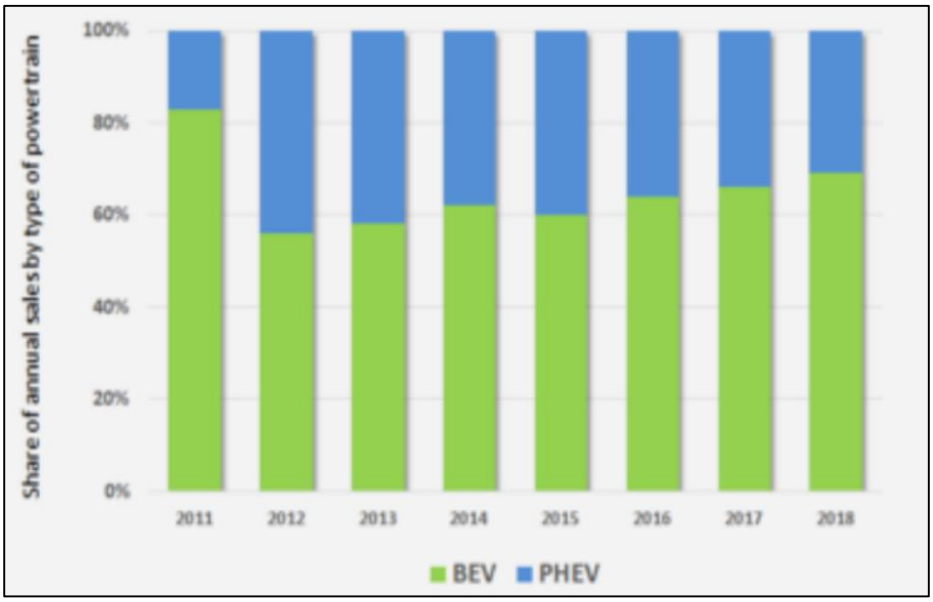


Fig 2: Global sales of electric vehicle

The first law of thermodynamics, also known as the law of conservation of energy, states that energy cannot be created or destroyed, it can only be transformed from one form to another. Solar energy is the energy of the sun is a huge source of energy in the form of heat and light produced by nuclear fusion in its core. Nuclear reactions release energy transfer moving towards the surface of the sun. Earth absorbs approximately 3,850,000 exajoules of solar energy per year and the fundamental energy from it is light energy. Of the heat and light, some systems use thermal energy for heating and relaxation, while others convert or convert light into electrical energy. In the current situation, the energy crisis is an important and urgent issue that needs to be resolved. It's basically checking the availability of to increase generation or to mount more generators. In addition, systematic demand-side management plans such as demand control and shift-or-reduce methods (DSM) 4244 should be adopted. In the current scenario, fossil fuel use for transportation is recommended. Inland transportation is increasing rapidly and presents a threat in the near future.

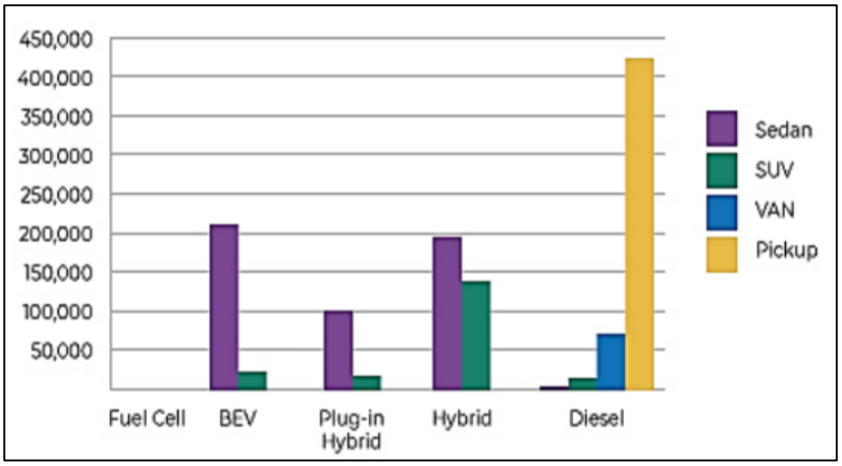


Fig 3: Sales of different E-vehicles

Harms of practicing means and these vehicles lead to environmental pollution and health problems related to the living environment. In line with this the upcoming mode of transport is E-vehicles keeping concern about the environment. The disadvantage of this mode of transport is the use of electric power for propulsion and the switch to transport mode will place additional load on the existing vehicle infrastructure and configuration. The advent of power electronics has made E-vehicle applications (EVs) smart and green among other transportation systems.

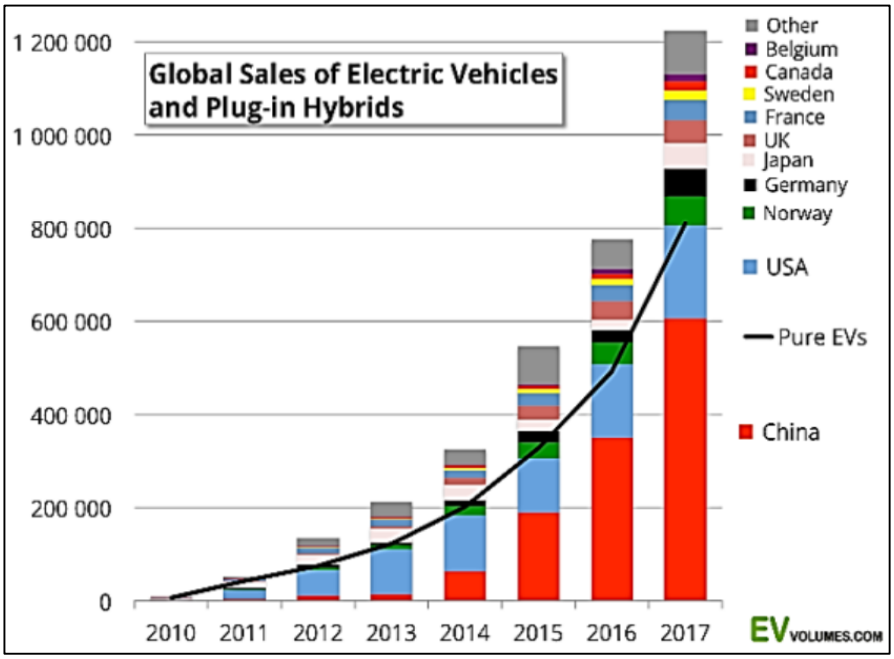


Fig 4: Global sale of E-Vehicles

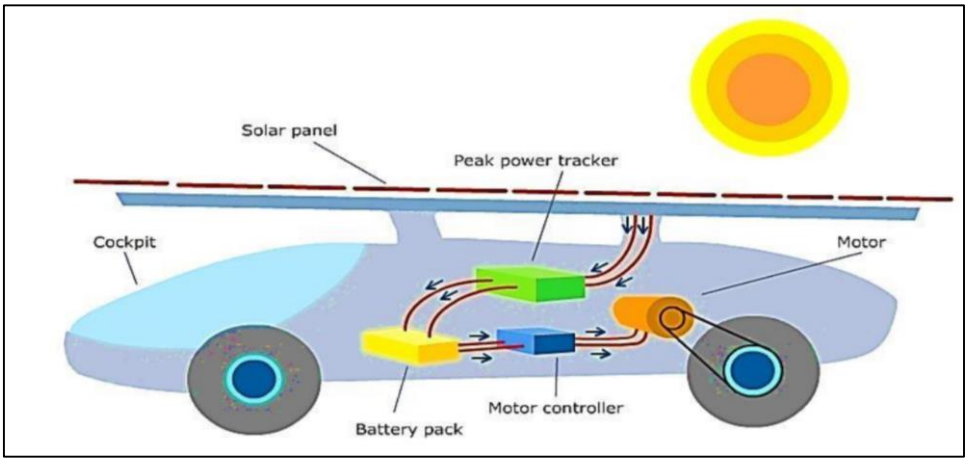


Fig 5: Working of the E-vehicle

Electronic charging controller and solar panel helps to charge and control the battery power in automotive operation with proper protection of output depth and overcharging etc. The 140 Wp solar module is used with output from 24V to 25V DC standard test conditions (STC) are present in the sample sample. Batteries are fully charged before use. Solar panel installed charger after converting light into electricity. The integrated features maintain a good setting as the availability of solar insolation is as important as the location of the car. Typically, in India, between 9:30 am and 3:30 pm with proper solar panel installation the best power output can be achieved. The motor installed in the car can perform the intended movement as in conjunction with a powerful battery controller.

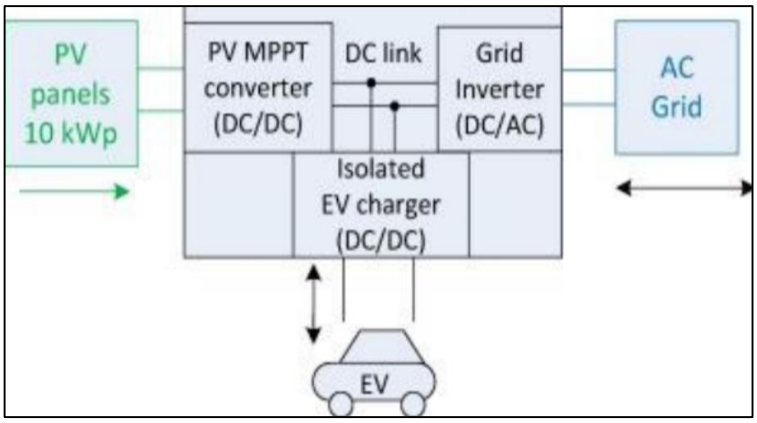


Fig 6: Functional block diagram

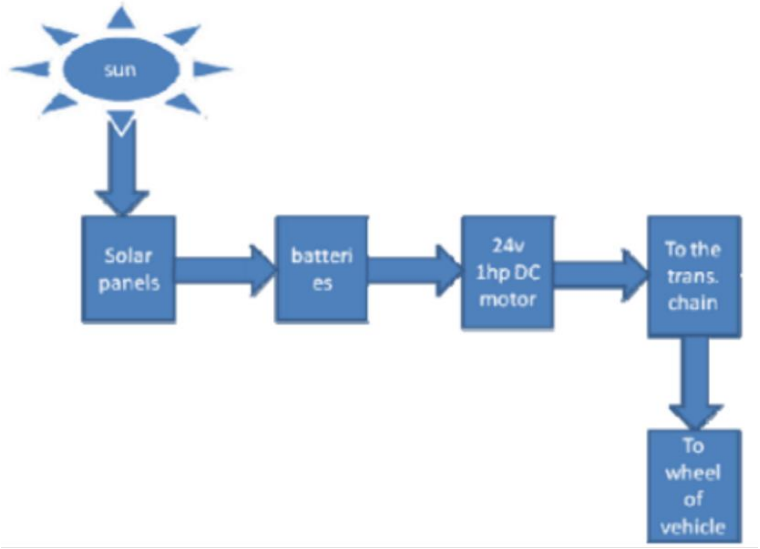


Fig 7: Block diagram solar E-vehicle

Components used	Range	Quantity
Batteries(heavy inverter batteries)	24V 190Ah	2*12V
Solar module	140Wp(Watt Peak)	1
Connecting Cables	Motor connection:-25Sq.mm high voltage cables. Solar module to charge controller unit:-1Sq.mm Charge controller to battery unit:-2Sq.mm	10 meters 1 meter 1 meter
Motor	High torque DC motor 1Hp=746W	1

Fig 8: The important parts of the E-vehicle

Battery
The battery is a very important component and can be charged electrically in a special power station. Battery autonomy is an important design limit. In addition to the standard technology of lithium ion batteries, there are at least four major battery technologies now represented: Solid state, aluminum ion, lithium sulfur, and air metal. Battery-based EVs provide a 200-mile run with full charge and usually a distance of 70–125 miles looks very normal.

Motor Controller
Vehicle controls are electronic circuits that control the operation of the engine as required by the road conditions used by the user. The controls are built with enough precision to feed the engine and take adequate safety. Power consumption adjustment is a special function of reducing consumption and improving battery autonomy.

Electric Engine
Electric motors can be programmed to use current or direct current power (AC or DC) to move forward with suitable motors. AC motors are usually less expensive and easier than DC motors which typically show 82% to 96% efficiency compared to those DC motors reaching approximately 95% - 96% efficiency when fully charged. AC motors are more common due to their less maintenance and cost compared to DC motors.

DC/DC Converter
An optical power connector capable of setting the DC output as required. The pulsating DC input may appear on a simple converter or battery. A certain level of output voltage is controlled by changing the elements in the converter. For example, a 48V / 12V module, 100W that converts a 48V battery to a 12V battery in loads such as lights, signals and wipers, etc.

Peak Power Tracker
The Maximum power point tracker (MPPT) is a machine that attempts to increase the efficiency of solar systems by setting the appropriate operating environment to output and give maximum effect to subsequent connected systems. It can be a microprocessor-controlled DC / DC step down converter controlled by a microprocessor used by the solar power system to charge the battery.

Conclusion
The present mix mode of transportation is having adverse emissions to the atmosphere and the current practice to reduce the impacts of these emissions is of greater importance. The practice of electric vehicles (EV) in this regard added greater value and provided large scope to the development and research. The associated government policies and benefits to both the manufacturers and users need to be revisited to enhance the E-vehicle market. But the popularity of silent and eco-friendly electric vehicles is increasing now a day's creating bundle of opportunities for the enterprises and users.

References

1. Cole J. Nissan Passes 50.000 LEAFs Sold Worldwide, Inside EVs, 2013. [updated 14th February 2013, 20th February 2013].
2. Lott MC, Trigg T. Electric vehicle deployment-Where should we be today, 2013.
3. Snider A. Pentagon Places Big Bet on Vehicle to Grid Technology, Greenwire, 2013.
4. Musti S, Kockelman KM. Evolution of the household vehicle fleet: Anticipating fleet composition, PHEV adoption and GHG emissions in Austin, Texas, Transportation Research Part A: Policy and Practice. 2011; 45(8):707-720.
5. Carroll S. The smart move case studies, The Centre of Excellence for Low Carbon and Fuel Cell Technologies (CENEX),
6. G Carli, SS Williamson. Technical Considerations on Power Conversion for Electric and Plug-in Hybrid Electric Vehicle Battery Charging in Photovoltaic Installations, IEEE Trans. Power Electron, 2011; 28(12):5784-5792.
7. Robert Foster, Maid Ghassemi, Alma Cota. Solar energy renewable energy and the Environment, Posts & Telecom Press, 2010, 2-3.
8. FU Zinian. Solar thermal utilization, China Science and Technology University Press, 2014, 273.
9. CC Chan. Electric hybrid and fuel-cell vehicles: architectures and modeling, IEEE Transactions on Vehicular Technology. 2012; 59(2):589-598.
10. K Rajashekara. Present status and future trends in electric vehicle propulsion Technologies, IEE E J. Emerging Sel. Topics Power Electron, 2014, 1

Proceedings of the International Conference on Sustainable Computing and Data Communication Systems (ICSCDS-2022)
IEEE Xplore Part Number: CFP22AZ5-ART; ISBN: 978-1-6654-7884-7

Solar Flare Prediction using Machine Learning Algorithms on RHESSI Dataset

Moulshree
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
moulshree_2k18co211@dtu.ac.in

Munira Anjum
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
muniraanjum_2k18co215@dtu.ac.in

Anurag Goel
Department of Computer Science & Engineering
Delhi Technological University
Delhi, India
anurag@dtu.ac.in

2022 International Conference on Sustainable Computing and Data Communication Systems (ICSCDS) | 978-1-6654-7884-7/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICSCDS53736.2022.9760755

***Abstract* — Machine learning has been in the industry for a long time and has been applied in diverse sectors, including traffic control, self-automated automobiles, recommender systems, security control, space exploration, etc. In the space exploration domain, the predictions made by machine learning and deep learning approaches using the videos captured by observatories help discover new theories about space. Solar Flare prediction is one of the machine learning applications. Solar flare prediction is complex, and the exact reason behind these massive eruptions is still a mystery. Solar flares are detectable through a telescope, and the particles scattered after the eruptions can reach the Earth within eight minutes. The frequent disturbances in the active spot of the sun can result in numerous disruptions on Earth, including power outages, disruptions in communication systems, etc. Therefore, it becomes crucial to have preliminary information on solar flares so that timely preparation can be done to combat their effects on Earth. In this work, we analyze the various machine learning techniques and deep learning approaches for solar flares prediction using the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) dataset.**

Keywords- Solar Flares, Solar Flares Prediction, Machine Learning, Long Short-Term Memory, Neural Network, Convolutional Neural Network

I. INTRODUCTION

Solar flares are an eruptive phenomenon of the sun. It can go under chemical reactions and release a tremendous amount of energy, which can harm Earth's atmosphere. The solar flares may result in disturbances, such as coronal mass ejections that erupt from Earth's active spots, which are detrimental. The particles that emerge from these solar flare eruptions can reach the Earth within a few minutes. Although these harmful radiations cannot penetrate the Earth's atmosphere and harm any humans on the ground, they can affect Earth's communication and power infrastructures. Solar flares have been a cause of various outrages, including the adverse impact on the revolving satellites, harm to the astronauts working in outer space, and damaging consequences on aviation and power grids.

There have been instances when a solar flare has hit Earth and caused massive damage. A recent reported incident is when a solar flare entered Earth's magnetic field on 31st October 2021, however, it was not catastrophic, but it did lead to the creation of auroras in the night sky at some specific locations. A solar storm was reported in May 1967. The telecommunication between the United States of America and the United Kingdom was jammed, and the radars were out of service. This was all due to a massive solar flare. Later in March 1989, North America faced 12 hours of blackout when a powerful coral mass ejection affected over 200 power grids in Quebec, Canada, by causing electric currents in the ground. We have witnessed harmless solar flare hits, but there have also been incidents where the communication on Earth was damaged, leading to monetary losses. Learning from the previous encounters, it is paramount to predict solar flares and be prepared before any such calamity strikes Earth.

There are immeasurable radiations that are roaming in space. In these radiations, the particles emerging from solar flare eruptions speed up and accelerate close to the speed of light. The strong solar radiations emit energetic protons that can damage electronic circuits present inside the satellite and even affect the biological DNA of the astronauts [6, 8]. Space enthusiasts, especially those flying over the north pole, can get exposed to these radiations. Humans have been studying the solar phenomenon for centuries and studying sunspots. The sun has an eleven-year cycle for sunspots. As per the research, by 2025, the sun will reach its solar maximum, and it will be one of the most significant solar activities ever to take place.

In this work, we analyze the various machine learning techniques [12] and deep learning approaches for solar flares prediction, namely Logistic Regression, Random Forest, Decision Tree, Artificial Neural Network (ANN), Long Short-Term Memory (LSTM), and Convolutional Neural Network (CNN). The performance is evaluated using the Reuven Ramaty High Energy Solar Spectroscopic (RHESSI) dataset [25], which has recorded numerous X-ray and gamma-emission events over more than an eleven-year solar cycle.

In this paper, the structure followed is – Section II details the related work, and Section III describes the various machine learning approaches used. The implementation is discussed in Section IV, followed by the analysis of results in Section V while Section VI concludes the paper

II. RELATED WORK

The need for prior prediction of solar flares is critical since the space phenomenon becomes more and more uncertain. Many researchers have attempted to predict the mass ejection from the sun in the form of solar flares by using various machine learning and deep learning techniques. In [1], Naoto Nishizuka et. al. attempted to predict the solar flares using a

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

Deep Flare net and worked towards developing an operational solar flare prediction model and categorized them into parts. T. Colak et. al. [2] used features like daily sunspots and McIntosh classification [11] as input to their solar flare prediction algorithm. The algorithm is built on two neural networks. It was known that larger sunspots with a more complex flex structure are bound to result in larger flares [20] and have a high chance of repeated flares that are present in the same active regions.

Various machine learning approaches are used for solar flares prediction, such as decision tree and learning vector quantization [21], multilinear regression model, and the logistic regression model [22]. Some other approaches include a combination of Support Vector Machine (SVM) [26] and K-Nearest Neighbour (K-NN) [23], combining K-Nearest Neighbour (K-NN) with nearest centroid algorithm, randomized trees [5], Support Vector Machine (SVM) combined with regression [24] and an ensemble of four classifiers [7, 9].

Another approach used for this purpose is a multivariate time series decision tree where the cluster potential of time series analysis was studied based on the flaring activities of the sun [13]. While doing the background studies and reading a similar research paper, we found that the RHESSI data is sometimes hard to study. This makes solar flares detection tough, and this was too discussed by Angryk [14]. The solar flares are divided into categories based on the emissions. These categories are namely B, C, M, and Q classes. Each flare is greater than the previous flare and has a greater impact than the previous one, that means the C class is much stronger than the B class. Here for the sake of the experiment. They were studying the C, B, Q as negative classes for flares and M class as a positive class. Due to this, those classes were tagged as non-flaring events. In the study of [15-18], it was noticed that marking a class as positive or negative is not an efficient method and C classes are valuable when studying the sunspots. Thus, before predicting the sunspots, it was paramount to also learn about the types of sunspots [19].

M-solar flares and coronal mass ejections affect Earth the most. While we mark the steps of this classification and implement the model, we can in terms, see how the prediction of flares can be accomplished through machine learning techniques [3, 4, 10]. To make the model, we need a data-driven approach as in earlier times, people did not believe in the data's time-series nature. We read about the time series forest, which tried to predict the forecasts and the flares while not minimizing the number of false negatives. Through this method which came out in 2020, they were able to find non-flaring activities. The studies have worked effectively to improve the solar flare prediction methods and build a more promising model. In [27], the intensity of solar flares are predicted using Long Short-Term Memory (LSTM) model.

III. MACHINE LEARNING APPROACHES

The various traditional machine learning approaches namely Logistic Regression, Decision Tree Classifier and Random Forest Classifier along with the neural network architectures namely Artificial Neural Network (ANN), Long Short-term Memory (LSTM), and Convolutional Neural Network (CNN) are considered for comparative analysis of solar flare prediction.

1) Logistic Regression

Logistic Regression is highly used and one of the most popular supervised learning techniques. Although it has a list of applications, it is highly used to predict a categorical dependent variable with the help of a given set of independent variables, providing its probabilistic value. When it comes to finding solutions for classification problems, Logistics Regressions come in very handy. Since solar flares prediction is too a type of classification problem, logistic Regression was used.

2) Decision Tree Classifier

A Decision Tree is mostly preferred mostly for solving classification problems. As for its structure, it is a tree-structured classifier, where the internal nodes of the tree represent the dataset features, branches represent decision rules, and the outcomes are represented by the leaf nodes. In simpler words, it is a graphical representation explaining the process of getting different outcomes for a given problem based on given conditions. It provides a simpler understanding of its logic as it mimics human thinking ability while making decisions. There can be the possibility of overfitting in this classifier as well as complex issues as the number of layers keeps increasing.

3) Random Forest Classifier

For the next algorithm, we more to a classifier that makes use of ensemble learning – Random Forest. This makes use of a variety of classifiers and combines them to further better the model's performance. Moving a step ahead of the decision tree, it contains a number of decision trees that are applied to subsets on the dataset. It takes their average to improve the prediction accuracy. This technique requires less time for training than other algorithms and provides high accuracy in prediction, even if the dataset is large or has missing portions.

4) Artificial Neural Network (ANN)

Artificial Neural Network (ANN) attempts to mimic the working of the human brain and the network of its neurons so that our computers have the option of not only understanding things the way humans do but also make decisions in a human-like manner. ANNs provide numerous advantages like storing data on the entire network rather than on a database. After an ANN has been trained, the information provided may possibly produce an output even with inadequate data and the extortion of one or more ANN cells fail in averting the process of generating output.

An Artificial Neural Network (ANN) is an integration of various layers where each layer has its own different function. These layers are - namely, the input layer that accepts the input features, the output layer that provides the final prediction, and a bunch of hidden layers that are responsible for learning the weights. While training, the input data is passed from the input layer to the hidden layers, where the weights corresponding to each input feature (attribute) is applied and finally reached to the output layer where the final output is generated after applying some kind of activation function, for example, Sigmoid or ReLU activation function. The computation of output error is done at the output layer, which is then backpropagated through the hidden layers to update the weights.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

5) Convolutional Neural Network (CNN)

Convolutional Neural Network is a well-trusted and studied neural network that is known for working with data that can be represented in imagery format. A convolution network is an integration of various layers, each defining a different function. When the data passes through these layers, it transforms depending on the functions. CNN's are known to be specialized in image classification problems. This is because CNN's are very effective in reducing the number of parameters without the quality loss of the data. This turns out to be very helpful when we are using imagery data as images have high dimensionality.

There are two main parts to the architecture of a CNN. First is the process called Feature Extraction, a convolutional tool to separate and identify various features of the imagery data for its analysis. Second is the process in which a fully connected layer utilizes the output and the extracted features from the previous convolution process and then helps predict the class of the image. CNN consists of three types of layers: convolutional, pooling, and fully connected layers.

6) Long Short-Term Memory (LSTM)

Long Short-Term Memory (LSTM) is a kind of Recurrent Neural Network (RNN) that tackles the problem of long-term dependencies. LSTM is well-knowing for preserving the information over a long period and is hence used for classification problems for time-series data and complex sequence prediction problems

LSTM has a chain-like structure with different memory blocks called cells. These memory blocks have the purpose of retaining information and remembering things. [1] The mandated changes in these memories are processed by making use of three types of gates- Forget Gate, Input Gate and Output Gate. The Forget gate removes the unnecessary information from the LSTM that helps optimize the network's performance. Using a three-step process, the Input gate adds the information to the cell state, and the Output gate performs the job of picking the meaningful knowledge from the current cell state, also through a three-step process, and displaying it as the output.

IV. IMPLEMENTATION

After doing the background research and discussing the algorithms used for the comparative study, we dive deeper into how the implementations were carried.

A) Experimental Setup

The experiments were carried out on a 64 bit Intel Core i7-8750H CPU at 2.20 GHz, 16 GB RAM running Windows 10. Python programming language is used for the implementation.

B) Dataset Used

Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) [2] is a mission that focuses on the sun and was launched by the National Aeronautics and Space Administration (NASA) to monitor the activities on the sun specifically exploring the particle acceleration and energy burnouts when the solar flares occur. The RHESSI dataset has a recording of over 1,20,000 X-ray events, 27 with gamma-ray line emission and 42 with gamma-ray emission above 300 keV. This data covers more than a complete 11-year solar cycle.

C) Data Preprocessing

The raw data file is in Flexible Image Transport System (FITS) format. FITS standards were specifically designed for astronomical data, which makes it the most commonly used digital file format in astronomical data research. To ease the implementation, the input data which is in FITS format, has been converted to Comma-Separated Values (CSV) format. The train-test split is done using the standard 80-20 ratio, which means 80% of the samples are included in the training data, and the rest 20% of samples are used as testing data.

D) Data Visualization

We have also analyzed the most influential features in solar flare data. For a given solar flare data sample, three features largely influence the determination of the energy range of the respective solar flare – Duration, Peak Intensity, and Total Counts. These features influence each other and influence the remaining features as well. We have plotted the relations of these features with the rest of the features in the input data for better interpretation.

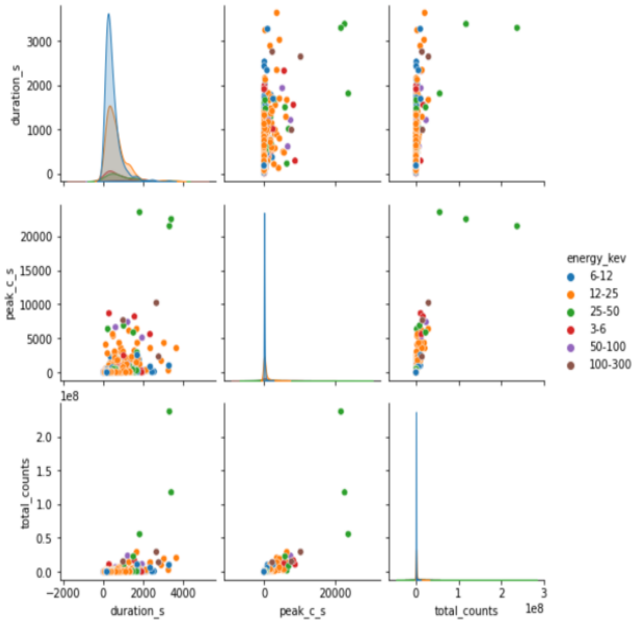


Fig. 1. Impact of Duration, Peak Intensity and Total Counts of Solar Flares on their Energy Range

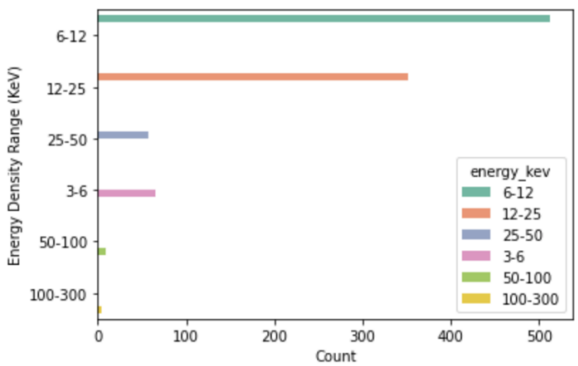


Fig. 2. Change in Energy Density Range and Total Count of Solar Flares

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

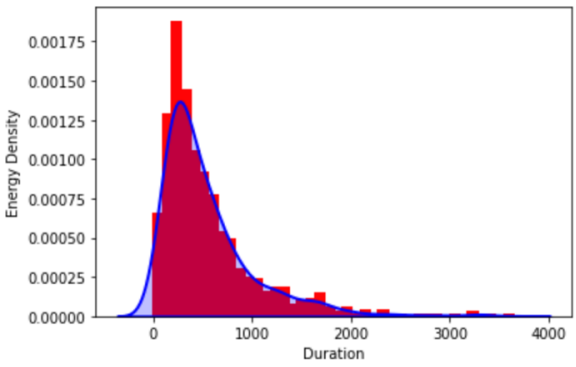


Fig. 3. Change in Energy Density with respect to Duration of Solar Flares

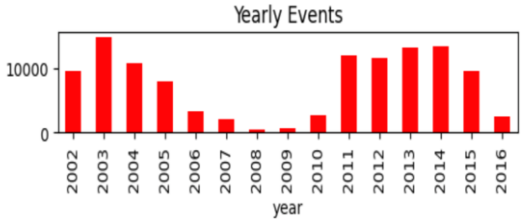


Fig. 4. Solar Flares Events during 2002 - 2016

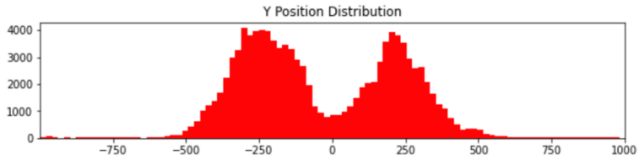


Fig. 5. Intensity Range of Solar Flares during 2002 – 2016 (in Histogram format)

Figure 1 shows the influence of Duration, Peak Intensity, and Total Counts on the energy range of the solar flare. The size of 1000 is taken because of the assumption that whatever pattern is going to be displayed in those 1000 flares, they will be continued over for the rest of the dataset as well. From Figure 2, it can be observed that the majority of the solar flares are in the 6-12 KeV energy range, followed by the solar flares in the 12-25 KeV energy range. The density plots in Figure 1 provide a more balanced view of the relationships between the three features, namely Duration, Peak Intensity, and Total Counts. The first density plot in Figure 1 reveals that the solar flares in the energy range of 6-12 KeV last for the longest time, followed by the solar flares in the 12-23 KeV range. Therefore, we can infer that the solar flares having energy ranges in 6-12 KeV and 12-25 KeV usually occur in a cluster and hence have high values for total_counts. The second density plot in Figure 1 shows the intensity distribution of solar flares over the different energy ranges and it can be infer that the solar flares that are in the 12-25 KeV range have the highest intensity since the solar flares peaked to a specific energy numerous times in just a second. Figure 2 portrays the relationship between the Energy density and count of solar flares and Figure 3 portrays the relationship between the Energy density and duration of solar flares.

We have also studied the occurrences of the solar flares events from 2002 to 2016 to identify any kind of pattern. Figure 4 shows the yearly counts of the solar flares events from 2002 to 2016. This pointed at the symmetric property in trends. The solar flares intensity is also plotted from 2002 to 2016 in histogram format in Figure 5.

E) Evaluation Metrics

We have used the several standard metrics used for classification namely Precision, Recall, F1-Score, Accuracy and True Skill Statistic. These metrics are defined over the True Positives (TP) that are the positive class samples that are correctly interpreted by the model, True Negatives (TN) which are the negative class samples that are correctly predicted by the model, False Positives (FP) represent those samples that actually belongs to the negative class but are incorrectly predicted by the model as positive and False Negatives (FN) which represent those samples that actually belongs to the positive class but are incorrectly predicted by the model as negative.

1. Precision

Precision is defined as the ability of a model to correctly classify the positive values.

$$\text{Precision} = \frac{TP}{TP+FP}$$

2. Recall

Recall can be defined as the number of positive values correctly predicted out of the total positive values.

$$\text{Recall} = \frac{TP}{TP+FN}$$

3. F1 Score

F1 score is important in determining the accuracy and precision of our classifiers. This score is the harmonic mean of Precision and Recall. It also showcases the robustness of the model. The greater the value of the F1- score the higher is the efficiency of the defect prediction model.

$$\text{F1 Score} = \frac{2*Precision*Recall}{precision+Recall}$$

4. Accuracy

Accuracy helps us to quantify the rightly made decisions by the model.

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

5. True Skill Statistic (TSS)

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

TSS is an extremely important metric for solar flare prediction. TSS represents irregularities between the observations and predictions in the confusion matrix. TSS takes omission and commission errors into account along with success. TSS value ranges from -1 to +1 where +1 indicates a perfect agreement and a value less than 0 indicates a model that is not better than random.TSS value in the range of 0.4 to 0.8 is desirable value. TSS value converges to 1 when the False-Negative rate is minimized. The best features of TSS are that prevalences and the size of the validation set do not influence it.

$$TSS = TRP + TNR - 1$$

Where $TPR = \frac{TP}{TP+FN}$

And $TNR = \frac{TN}{TN+FP}$

[TP = True Positive
FP = False Positive
TN = True Negative
TN = True Negative
TPR = True Positivity Rate (Sensitivity)
TNR = True Negativity Rate (Specificity)]

V. RESULTS ANALYSIS

The predictions for the various machine learning algorithms are shown in Table 1. The results show that CNN has shown the best performance while Linear Regression has shown the worst performance on the RHESSI dataset. Random Forest has also shown decent performance and has performed best in the traditional machine learning approaches.

Table 1. Results

Algorithm	Prediction Results				
	Precision	**Recall**	**F1-Score**	**Accuracy**	**TSS**
Linear Regression	0.75	0.79	0.73	0.78	0.61
Decision Tree	0.82	0.81	0.80	0.82	0.68
Random Forest	0.83	0.85	0.82	0.84	0.76
Artificial Neural Network (ANN)	0.81	0.81	0.80	0.81	0.74
Convolutional Neural Network (CNN)	**0.86**	**0.861**	**0.83**	**0.86**	**0.77**
Long Short Term Memory (LSTM)	0.82	0.81	0.81	0.81	0.75

Linear Regression gives True Skill Statistic (TSS) value as 0.61 and precision as 0.75, which is the lowest. The Decision tree classifier and Random forest results are also comparable, with precision being 0.82 and 0.83, respectively; however, F1 score and TSS values have a significant difference for both the models. The F1 score of Decision tree is 0.81, while the F1 score of Random forest is 0.85. The TSS of Decision tree is 0.68 while the TSS of Random forest is 0.76. We observe that for an algorithm to be efficient, it does not have to perform well in only one parameter. On the contrary, LSTM and Artificial Neural Networks had comparable results in all parameters differing by only a 0.1 value. The two algorithms resulted in the same recall.

VI. CONCLUSION

In this work, we have explored and visualized the Reuven Ramaty High Energy Solar Spectroscopic Imager (RHESSI) dataset to identify the patterns of the solar flares. The high volume of observational data can be utilized further to discover new flare-sensitive features and potentially enhance the automated operational forecasts which are still mainly done by human experts. In this work, the various machine learning approaches, namely Linear Regression, Decision Tree, Random Forest, Artificial Neural Network (ANN), Convolutional Neural Network (CNN), and Long Short Term Memory (LSTM), are implemented for solar flares prediction using the RHESSI dataset. From the results, it has been observed that Convolutional Neural Network has shown the best results while Linear Regression shows the worst performance. For future work, Synthetic Minority Over-sampling Technique (SMOTE) or Time Series Classifier can be implemented to improve the prediction results.

REFERENCES

1. Nishizuka, Naoto, Yûki Kubo, Komei Sugiura, Mitsue Den, and Mamoru Ishii. "Operational solar flare prediction model using Deep Flare Net." Earth, Planets and Space 73, no. 1 (2021): 1-12.
2. Ahmed, Omar W., Rami Qahwaji, Tufan Colak, Paul A. Higgins, Peter T. Gallagher, and D. Shaun Bloomfield. "Solar flare prediction using advanced feature extraction, machine learning, and feature selection." Solar Physics 283, no. 1 (2013): 157-175.
3. Bobra, Monica G., and Sebastien Couvidat. "Solar flare prediction using SDO/HMI vector magnetic field data with a machine-learning algorithm." The Astrophysical Journal 798, no. 2 (2015): 135.
4. Wang, Naiyan, and Dit-Yan Yeung. "Learning a deep compact image representation for visual tracking." Advances in neural information processing systems 26 (2013).
5. Liu, Chang, Na Deng, Jason TL Wang, and Haimin Wang. "Predicting solar flares using SDO/HMI vector magnetic data products and the random forest algorithm." The Astrophysical Journal 843, no. 2 (2017): 104.
6. Eastwood, J. P., E. Biffis, M. A. Hapgood, L. Green, M. M. Bisi, R. D. Bentley, Robert Wicks, L-A. McKinnell, M. Gibbs, and C. Burnett. "The economic impact of space weather: Where do we stand?." Risk Analysis 37, no. 2 (2017): 206-218.
7. Benvenuto, Federico, Michele Piana, Cristina Campi, and Anna Maria Massone. "A hybrid supervised/unsupervised machine learning approach to solar flare prediction." The Astrophysical Journal 853, no. 1 (2018): 90.
8. Leka, K. D., and G. Barnes. "Photospheric magnetic field properties of flaring versus flare-quiet active regions. II. Discriminant analysis." The Astrophysical Journal 595, no. 2 (2003): 1296.
9. Georgoulis, Manolis K., D. Shaun Bloomfield, Michele Piana, Anna Maria Massone, Marco Soldati, Peter T. Gallagher, Etienne Pariat et al. "The flare likelihood and region eruption forecasting (FLARECAST) project: flare forecasting in the big data & machine

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

learning era." Journal of Space Weather and Space Climate 11 (2021): 39.

10. Wang, Xiantong, Yang Chen, Gabor Toth, Ward B. Manchester, Tamas I. Gombosi, Alfred O. Hero, Zhenbang Jiao, Hu Sun, Meng Jin, and Yang Liu. "Predicting solar flares with machine learning: Investigating solar cycle dependence." The Astrophysical Journal 895, no. 1 (2020): 3.
11. McIntosh, Patrick S. "The classification of sunspot groups." Solar Physics 125, no. 2 (1990): 251-267.
12. Choudhury, Sumouli, and Anirban Bhowal. "Comparative analysis of machine learning algorithms along with classifiers for network intrusion detection." In 2015 International Conference on Smart Technologies and Management for Computing, Communication, Controls, Energy and Materials (ICSTM), pp. 89-95. IEEE, 2015.
13. Bhattacharjee, Shamik, Rasha Alshehhi, Dattaraj B. Dhuri, and Shravan M. Hanasoge. "Supervised convolutional neural networks for classification of flaring and nonflaring active regions using line-of-sight magnetograms." The Astrophysical Journal 898, no. 2 (2020): 98.
14. Boubrahimi, Soukaina Filali, Berkay Aydin, Dustin Kempton, and Rafal Angryk. "Spatio-temporal interpolation methods for solar events metadata." In 2016 IEEE International Conference on Big Data (Big Data), pp. 3149-3157. IEEE, 2016.
15. Sarker, Toqi Tahamid, and Juan M. Banda. "Solar Event Tracking with Deep Regression Networks: A Proof of Concept Evaluation." In 2019 IEEE International Conference on Big Data (Big Data), pp. 4942-4949. IEEE, 2019.
16. Ma, Ruizhe, Azim Ahmadzadeh, Soukaina Filali Boubrahimi, and Rafal A. Angryk. "Segmented dynamic time warping: A comparative and applicational study." In Emerging Technologies and Applications in Data Processing and Management, pp. 1-19. IGI Global, 2019.
17. Huang, Xin, Huaning Wang, Long Xu, Jinfu Liu, Rong Li, and Xinghua Dai. "Deep learning based solar flare forecasting model. I. Results for line-of-sight magnetograms." The Astrophysical Journal 856, no. 1 (2018): 7.
18. Fernandez Borda, Roberto A., Pablo D. Mininni, Cristina H. Mandrini, Daniel O. Gómez, Otto H. Bauer, and Marta G. Rovira. "Automatic solar flare detection using neural network techniques." Solar Physics 206, no. 2 (2002): 347-357.
19. Dhuri, Dattaraj B., Shravan M. Hanasoge, and Mark CM Cheung. "Machine learning reveals systematic accumulation of electric current in lead-up to solar flares." Proceedings of the National Academy of Sciences 116, no. 23 (2019): 11141-11146.
20. Lee, J. Y., Y. J. Moon, K. S. Kim, Y. D. Park, and A. B. Fletcher. "Prediction of daily maximum x-ray flux using multilinear regression and autoregressive time-series methods." Journal of The Korean Astronomical Society 40, no. 4 (2007): 99-106.
21. Bishop, Christopher M., and Nasser M. Nasrabadi. Pattern recognition and machine learning. Vol. 4, no. 4. New York: springer, 2006.
22. Boucheron, Laura E., Amani Al-Ghraibah, and RT James McAteer. "Prediction of solar flare size and time-to-flare using support vector machine regression." The Astrophysical Journal 812, no. 1 (2015): 51.
23. Guerra, Jordan Armando, A. Pulkkinen, and Vadim M. Uritsky. "Ensemble forecasting of major solar flares: First results." Space Weather 13, no. 10 (2015): 626-642.
24. Nishizuka, Naoto, Yûki Kubo, Komei Sugiura, Mitsue Den, and Mamoru Ishii. "Operational solar flare prediction model using Deep Flare Net." Earth, Planets and Space 73, no. 1 (2021): 1-12.
25. RHESSI Dataset: https://hesperia.gsfc.nasa.gov/rhessi2/
26. Florios, Kostas, Ioannis Kontogiannis, Sung-Hong Park, Jordan A. Guerra, Federico Benvenuto, D. Shaun Bloomfield, and Manolis K. Georgoulis. "Forecasting solar flares using magnetogram-based predictors and machine learning." Solar Physics 293, no. 2 (2018): 1-42.
27. Jiao, Zhenbang, Hu Sun, Xiantong Wang, Ward Manchester, Tamas Gombosi, Alfred Hero, and Yang Chen. "Solar flare intensity prediction with machine learning models." Space Weather 18, no. 7 (2020): e2020SW002440.

Authorized licensed use limited to: DELHI TECHNICAL UNIV. Downloaded on May 27,2022 at 10:03:39 UTC from IEEE Xplore. Restrictions apply.

Journal of Electronic Materials
https://doi.org/10.1007/s11664-022-09648-x

ORIGINAL RESEARCH ARTICLE

Spectroscopic Studies of Eu^{3+} Ion-Doped Antimony-Lead-Oxyfluoroborate Glasses for Visible Red Photonic Device Applications

D. Shanmukha Sai[2] · K. Swapna[1] · K. Deepthi[2] · J. V. Shanmukh Kumar[2] · Sk. Mahamuda[1] · M. Venkateswarlu[1] · P. Sruthi[2] · D. Amer[2] · A. S. Rao[3]

Received: 1 July 2021 / Accepted: 8 April 2022
© The Minerals, Metals & Materials Society 2022

Abstract

Antimony-lead-oxyfluoroborate (SbPbOFB) glasses doped with various concentrations of Eu^{3+} ions were prepared by employing melt quenching and their luminescence activity was characterized using different spectroscopic methods such as optical absorption, photoluminescence (PL) excitation, PL emission, and PL decay. The Raman spectrum was recorded for an un-doped SbPbOFB glass to identify various functional groups present. The optical absorption spectra recorded for Eu^{3+} ion-doped SbPbOFB glasses showed five bands corresponding to $^7F_0 \rightarrow {}^5L_6$, $^7F_0 \rightarrow {}^5D_2$, $^7F_1 \rightarrow {}^5D_1$, $^7F_0 \rightarrow {}^7F_6$ and $^7F_1 \rightarrow {}^7F_6$ transitions. The PL spectra measured under 394-nm excitation wavelength show five peaks corresponding to $^5D_0 \rightarrow {}^7F_0$, $^5D_0 \rightarrow {}^7F_1$, $^5D_0 \rightarrow {}^7F_2$, $^5D_0 \rightarrow {}^7F_3$ and $^5D_0 \rightarrow {}^7F_4$ transitions at 578, 592, 613, 652 and 701 nm, respectively. Among them, that pertaining to the $^5D_0 \rightarrow {}^7F_2$ transition appeared at 613 nm in the visible red region showing relatively high intensity. The Judd-Ofelt (J-O) parameters evaluated from the absorption spectral data are correlated with the prominent emission spectral features such as $^5D_0 \rightarrow {}^7F_1$ (592 nm), $^5D_0 \rightarrow {}^7F_2$ (613 nm) and $^5D_0 \rightarrow {}^7F_4$ (701 nm) to estimate radiative properties such as transition probabilities (A_R), branching ratios (β_R) and emission cross sections (σ_{se}). Quantum efficiency ($\eta\%$) of the prepared glasses was estimated by correlating the radiative lifetimes with the experimental lifetimes measured from the PL decay spectral features. The PL emission spectral features are used to estimate the CIE color coordinates (x, y) of the studied glasses. The measured β_R, σ_{se}, $\eta\%$, and the CIE coordinates estimated for the studied glasses revealed their suitability for visible red light-emitting device applications.

✉ K. Swapna
manchikalapudi80@gmail.com; kswapna80@gmail.com; swapnakon@kluniversity.in

✉ J. V. Shanmukh Kumar
shanmukh_fed@kluniversity.in

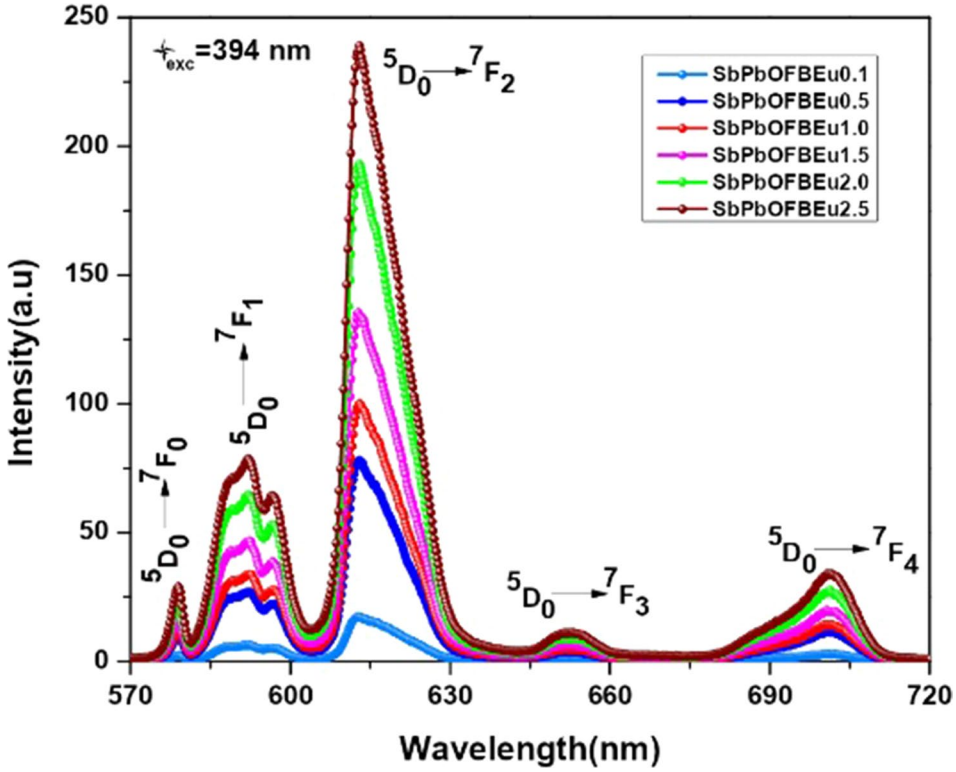
[1] Department of Physics, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, 522502 Guntur, Andhra Pradesh, India

[2] Department of Chemistry, Koneru Lakshmaiah Education Foundation, Green Fields, Vaddeswaram, Guntur, Andhra Pradesh 522502, India

[3] Department of Applied Physics, Delhi Technological University, Bawana Road, New Delhi 110 042, India

Published online: 10 May 2022

Graphical Abstract



Keywords Glasses · photoluminescence · raman spectrum · emission cross-sections · quantum efficiency · CIE coordinates

Introduction

Glassy materials doped with rare earth (RE) ions have been studied by many researchers around the globe to realize their widespread applications in diverse fields, such as lasers, light-emitting diodes (LEDs), optical switches and memory devices. Among various available glass matrices, the oxide-based glasses are found to be relatively more chemically stable host matrices.[1] Apart from the aforementioned scientific applications, oxide glasses also have a wide variety of applications in fabricating daily use materials such as lab glassware, optical instruments, x-ray protection devices, and fiberglass.[2] Most of the glass properties originate from the network formers that make the bulk mass of the glass. Properties of such bulk glasses can be substantially amended by means of network modifiers. These network-modifying agents exist only as single ions surrounded by cross-linked networks and reduce the relative number of strong bonds. This reduction results in lowering the melting point, altering thermal and electrical properties, and decreasing viscosity. This type of effect has a specific significance from the perspective of improving the chemical durability of the glass.[3]

Amongst the well-known glass hosts, borate glass materials are of great interest because of their good RE ion solubility, cost-effectiveness, moderately small melting point, high thermal stability, good transparency, and capability to form different structural units within the glass matrices.[4,5] Apart from such encouraging and interesting features, borate glass materials possess relatively high phonon energies, which is detrimental for their usage in photonic device applications. Borate glasses with relatively high phonon energies will incline more towards non-radiative transitions and reduce their quantum efficiency. However, such high phonon energies of borates can be controlled by adding heavy metal oxides (HMOs), namely, antimony (III) oxide (Sb_2O_3) to enhance the luminescence efficiency.[1]

Antimony-altered HMO glasses are very attractive because of properties such as extreme thermal expansion coefficient, high polarizability, good transparency, low dispersion, higher RE ion solubility, and good linear and non-linear refractive indexes. They are also very appropriate for the applications in power limiters, solid-state ionic gadgets, and broadband optical amplifiers at 1.5 μm.[6,7] Even though glass materials containing lead oxide (PbO) are categorized as toxic matter, PbO exists as a fine glass network

modifier. Moreover, PbO glasses have fascinating properties such as high dispersion and high refractive index.[8] Similarly, fluoride ions correspondingly are seen with the alteration of borate glass networks. Doweidar et al.[9,10] stated that F^- ions are apt to change the tetrahedral borate structures in the MF_2-B_2O_3 glass materials (M = Ba, Ca, Pb and Cd). Adding fluoride modifiers (NaF) significantly decreases the phonon energy of the glass matrix and results in decreasing non-radiative transitions. Such glasses executing radiative transitions dominating over non-radiative transitions are prone to have relatively high quantum efficiency when doped with RE ions.[11]

RE ions generate optical transitions within 4f-4f orbitals, which is critical for the evolution of optical devices such as color displays, LEDs, phosphor materials, and lasers. Both radiative and non-radiative processes occur within the same 4f levels. Glass materials doped with trivalent RE ions obtained much more recognition due to emissions within the optical region ranging from visible to infra-red (IR) as this is significant in laser emission.[12] Eu^{3+} ions are commonly used in several optical devices. Interestingly, europium has visible radiation (green or blue) in the two different oxidation states (green in Eu^{2+} and red in Eu^{3+}) under UV excitation. The Eu^{3+} ions possess well-known laser luminescence properties in the reddish-orange region with narrow emission bands and relatively higher lifetimes.[13]

Among the several RE ion-doped glasses, Eu^{3+} -doped borate glasses are mostly chosen for their red luminescence which has applications in making color displays as well as red LEDs.[14] The abovementioned various scientific merits of chemical components such as B_2O_3, Sb_2O_3, PbO, and NaF prompted us to prepare a new glass system, namely antimony-lead-oxy-fluoro-borate (SbPbOFB) glass, and doped it with Eu^{3+} ions particularly to recognize its usage in visible red light-emitting luminescent device applications.

Experimental

SbPbOFB glass materials doped with Eu^{3+} ions have been prepared by using rapid melt quenching. The composition details and their abbreviations used to represent the studied glasses are as given below:

SbPbOFBEu0.1: $70B_2O_3$ – 20PbO - $5Sb_2O_3$ - 4.9NaF – $0.1Eu_2O_3$
SbPbOFBEu0.5: $70B_2O_3$ - 20PbO - $5Sb_2O_3$ - 4.5NaF – $0.5\ Eu_2O_3$
SbPbOFBEu1.0: $70B_2O_3$ - 20PbO - $5Sb_2O_3$ - 4.0NaF – $1.0\ Eu_2O_3$
SbPbOFBEu1.5: $70B_2O_3$ - 20PbO - $5Sb_2O_3$ - 3.5NaF – $1.5\ Eu_2O_3$
SbPbOFBEu2.0: $70B_2O_3$ - 20PbO - $5Sb_2O_3$ - 3.0NaF – $2.0\ Eu_2O_3$
SbPbOFBEu2.5: $70B_2O_3$ - 20PbO - $5Sb_2O_3$ - 2.5NaF – $2.5\ Eu_2O_3$

Initially, the inorganic basic raw materials were properly weighed (6 g per individual set) according to their compositions as mentioned above and were finely ground in an agate mortar with a pestle. The well-ground ingredients were put into a high-grade silica crucible and heated in an electrical furnace at 800°C for 1 h. The homogeneous melt was quenched between the two preheated brass molds. After that, the glasses were annealed at 400°C for about 3 h to remove built-in thermal strains. The as-prepared glasses were polished well before undergoing characterization. The formed glasses were circular in shape and uniform in thickness (2.054 mm). The refractive indices of the glasses were measured by means of an Abbe refractometer. By using the Archimedes principle, employing water as an immersion liquid, the densities of the glasses were measured. The Raman spectrum for an un-doped SbPbOFB glass was recorded between 200 cm^{-1} and 1800 cm^{-1} using an Agiltron Inspire 2785 LE Raman spectrometer. The absorption spectra for the glasses was measured in the visible to IR regions between 200 nm and 2700 nm making use of a JASCO V-670 UV-vis-NIR spectrophotometer. PL excitation and emission spectra were recorded by means of a JASCO FP 8300 Spectrofluorophotometer. An Edinburgh FLSP900 spectrophotometer with a spectral resolution of about 0.1 nm was used to record time-resolved PL (TR-PL) spectra, in which a xenon flash lamp was used as an excitation source. All the measurements were done at ambient temperature.

Results and Discussion

XRD and Physical Properties Analysis

Figure 1 shows the XRD spectra recorded for un-doped and Eu^{3+} ion-doped SbPbOFB glasses. The un-doped and doped glasses show no sharp peaks confirming the glassy nature of the studied glasses. The addition of Eu^{3+} ions need not show any shifting in the XRD pattern of the glasses. Using the experimentally measured refractive index and density values, various other dependent parameters have been calculated for the glasses using the suitable mathematical expressions given in the literature and are shown in Table I.[15] An increase in the refractive index of the glasses doped with RE ions increases the polarizability of the glass. With the increase in density, the compactness of glass increases. An increase in average molecular weight increases the mass of the glass. With the increase in Eu^{3+} ion concentration, the interionic

distance in the glasses decreased. The other parameter is the minimum value of molecular electronic polarizability states that ensure the stability of SbPbOFB glasses doped with Eu^{3+} ions. The molar refractivity values are examined as the total sum of cationic refraction to the oxygen ion, and these result in an increase of molar refractivity with an increase in the Eu^{3+} concentration but this was due to the bonding nature in the glass matrices.[16] The as-prepared glasses show low field strength values which support the best solubility of the doped RE ion in host glass. The increase in the optical basicity with the increase in Eu^{3+} ions shows the surge in the ability of the oxide ions to donate the negative charge to the glass medium. The optical basicity values increase with the increase in Eu^{3+} ions illustrating the strong covalency in cation-oxygen bonds in the host glass.

Raman Spectral Analysis

Figure 2 illustrates the Raman spectrum recorded for an un-doped SbPbOFB glass at ambient temperature in between 200-1800 cm^{-1}. As depicted in Fig. 2, the Raman spectrum shows eight bands at 290 cm^{-1}, 418 cm^{-1}, 482 cm^{-1}, 657 cm^{-1}, 932 cm^{-1}, 1141 cm^{-1}, 1308 cm^{-1}, 1522 cm^{-1}, and 1666 cm^{-1}. Among all the bands, a broad band observed at 1308 cm^{-1} is relatively more intense caused by metaborate

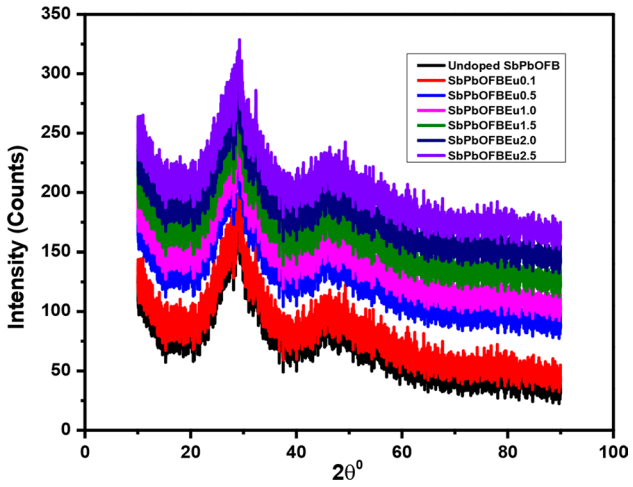


Fig. 1 XRD spectra of an un-doped SbPbOFB glasses and SbPbOFB glasses doped with different concentrations of Eu^{3+} ions.

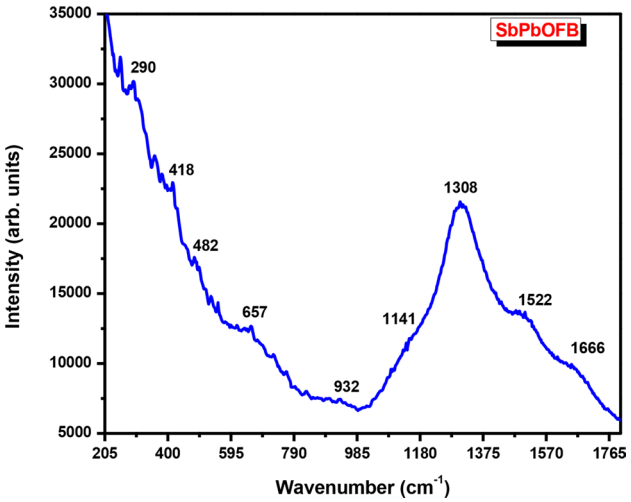


Fig. 2 Raman spectrum of undoped SbPbOFB glass.

Table I Physical Properties of Eu^{3+} ion-doped SbPbOFB glasses

Eu^{3+} concentration/physical properties	SbPbOFBEu0.1	SbPbOFBEu0.5	SbPbOFBEu1.0	SbPbOFBEu1.5	SbPbOFBEu2.0	SbPbOFBEu2.5
Refractive index (nd)	2.427	2.434	2.439	2.441	2.445	2.448
Density (*d*) (g/cc)	4.020	4.050	4.087	4.124	4.161	4.198
Average molecular weight (*M*) (g)	110.3	111.5	113.1	114.6	116.2	117.7
Eu^{3+} ion concentration ($\times 10^{22}$ ions/cm^3)	0.219	1.090	2.180	3.250	4.310	5.370
Mean atomic volume (g/cm^3/atom)	6.452	6.460	6.468	6.473	6.480	6.486
Optical dielectric constant (ϵ–1)	4.892	4.926	4.948	4.960	4.980	4.995
Dielectric constant (ϵ)	5.892	5.926	5.948	5.960	5.980	5.995
Reflection losses (*R*%)	0.173	0.174	0.175	0.175	0.176	0.176
Molar refraction (R_m) (cm^{-3})	17.01	17.12	17.23	17.32	17.43	17.52
Polaron radius (r_p) (Å)	3.100	1.820	1.440	1.260	1.150	1.070
Interionic distance (r_i) (Å)	7.700	4.510	3.580	3.130	2.850	2.650
Molecular electronic polarizability (α)($\times 10^{-23}$ cm^3)	0.084	0.169	8.480	5.680	4.280	3.440
Field strength ($\times 10^{15}$ cm^{-2})	3.120	9.100	14.40	18.80	22.70	26.30
Optical basicity (Λth)	0.419	0.420	0.421	0.422	0.425	0.424

(BO_3^-) triangular units linked to tetrahedral BO_4^- units.[17] The low-intensity peak at 290 cm^{-1} is mainly because of the combination of asymmetric stretching vibrations.[18,19] A Raman band observed at 418 cm^{-1} is due to bridged anion motion caused by symmetric stretching of the oxygen in the Pb-O-Pb units.[20] The middle bands observed at 482, 657, and 932 are a sign of pentaborate groups in the as-prepared borate glass system. The medium absorption peak at 1141 cm^{-1} is because of the B-O asymmetric stretching vibrations of tetrahedral BO_4 structures in the glass medium.[21]

Optical Absorption Spectral Studies

Figure 3 shows the optical absorption spectra recorded for different concentrations of Eu^{3+} ion-doped SbPbOFB glasses in visible and near infra-red regions, displaying several electronic transitions arising out of $^7F_0$ and $^7F_1$ as ground states to different higher excited states. In the case of europium ions, $^7F_0$ and $^7F_1$ levels act as ground states as the energy difference between them is about 360 cm^{-1}. Because of this small energy difference, at room temperature both energy states are populated thermally. As a result, Eu^{3+} ion-doped glasses show absorption bands coming from the ground state ($^7F_0$) as well as from the excited state ($^7F_1$).[22] The intensity of the optical absorption spectrum increases with an increase in Eu^{3+} ion concentration. Identification of absorption bands has been done in accordance with the results published by Carnal et al.[23] Fig. 3 shows five absorption bands corresponding to $^7F_0 \rightarrow {}^5L_6$, $^7F_0 \rightarrow {}^5D_2$, $^7F_1 \rightarrow {}^5D_1$, $^7F_0 \rightarrow {}^7F_6$, $^7F_1 \rightarrow {}^7F_6$ transitions observed at 393, 464, 533, 2091 and 2202 nm, respectively. Among all the absorption transitions observed, the $^7F_0 \rightarrow {}^7F_6$ transition that appears at 2091 nm in the NIR region is relatively more intense and is called a hypersensitive transition. The other transition, $^7F_0 \rightarrow {}^5D_2$, occurs because of the allowed electric dipole.[24] Optical absorption spectra play a significant role in determining the J-O parameters, i.e., Ω_2 and Ω_4 and Ω_6 by means of the J-O theory. Amongst those J-O parameters, Ω_2 relates to the structural environment around Eu^{3+} ions (short-range effect), whereas Ω_4 and Ω_6 parameters relate to the bulk properties such as viscosity and rigidity. In the Eu^{3+} ion-doped glass system, except for $^7F_1 \rightarrow {}^5D_4$ and $^7F_1 \rightarrow {}^5D_3$ transitions, the square reduced matrix element, i.e., $\|U^4\|$, was zero for other transitions.[25] In the present work, except for $^7F_0 \rightarrow {}^5L_6$ and $^7F_0 \rightarrow {}^5D_2$ transitions (of relatively less intensity), for the remaining transitions, the square reduced matrix elements, i.e., $\|U^2\|$ and $\|U^6\|$, are zero.[26] As a consequence, in the present work, the absorption spectral data are not sufficient to estimate J-O parameters by applying the J-O theory. For this reason, the necessary J-O parameters were evaluated by utilizing the emission spectral data.[27]

PL Excitation and Emission Spectral Analysis

Figure 4 shows the excitation spectra of SbPbOFB glasses doped with Eu^{3+} monitored at 613-nm emission. The excitation spectra shown in Fig. 5 display excitation bands corresponding to various transitions, namely, $^7F_0 \rightarrow {}^5D_4$, $^7F_0 \rightarrow {}^5G_2$, $^7F_0 \rightarrow {}^5L_6$, $^7F_0 \rightarrow {}^5D_3$, $^7F_0 \rightarrow {}^5D_2$, $^7F_1 \rightarrow {}^5D_1$, $^7F_0 \rightarrow {}^5D_1$, $^7F_0 \rightarrow {}^5D_0$, and $^7F_1 \rightarrow {}^5D_0$ at 361, 382, 394, 414, 464, 525, 532, 578, and 586 nm, respectively. The identification of excitation bands has been done in accordance with the results published by Carnal et al.[28] Among all the SbPbOFB glasses, SbPbOFBEu2.5 glass shows relatively more intense excitation peaks in comparison with the other glasses. Among all the excitation peaks observed, that pertaining to the $^7F_0 \rightarrow {}^5L_6$ transition (pure electronic transition) observed at 394 nm is relatively more intense and is employed while recording the PL emission spectra of the studied glasses.[13]

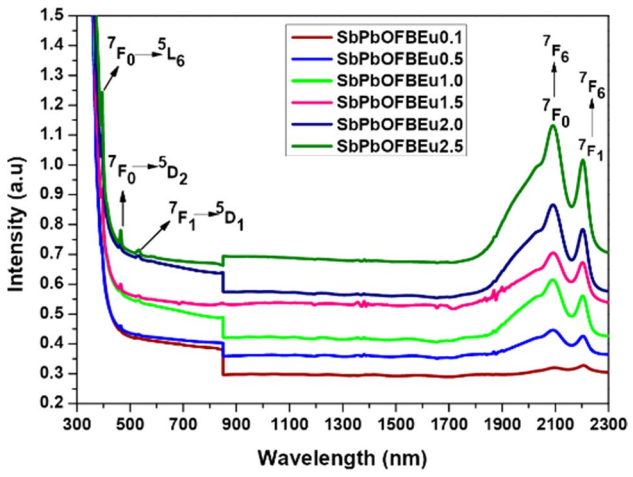


Fig. 3 Optical absorption spectra of Eu^{3+} ion-doped SbPbOFB glasses.

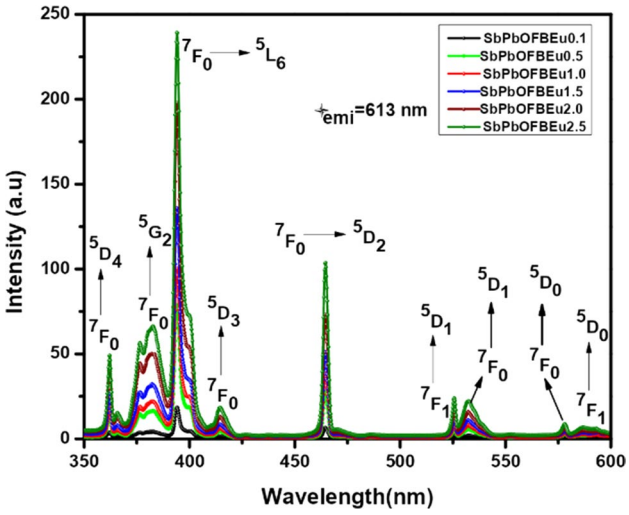


Fig. 4 Excitation spectra of Eu^{3+} ion-doped SbPbOFB glasses.

Figure 5 displays the emission spectra obtained for all the SbPbOFB glass samples doped with Eu^{3+} ions under 394-nm excitation wavelength. As shown in Fig. 6, the PL emission spectra consist of five emission peaks at 578, 592, 613, 652 and 701 nm correlated to the $^5D_0 \rightarrow {}^7F_J$ (J = 0,1,2,3,4) transitions, respectively. Amongst all emission transitions, the $^5D_0 \rightarrow {}^7F_2$ transition observed at 613 nm (visible red region) is the strongest of all the emissions. It is designated as induced or forced electric-dipole transition and can also be called a hypersensitive or pseudo-quadrupole transition. This transition looked like the admixture of the odd parity character into the $^4F_6$ wave function. This transition obeys the following selection rules $\Delta S = 0$ and $\Delta L \leq 2$ while $\Delta J \leq 2$.[29] Then, the strength of the hypersensitive transition was greatly strained by local symmetry about Eu^{3+} ions and the nature of surrounding ligands. The second strong and broad transition is at 592 nm which was present in the orange region is related to the magnetic dipole transition $^5D_0 \rightarrow {}^7F_2$ which is spin forbidden. The intensities of the emission bands are continuously increasing with Eu^{3+} ion concentration in the as-prepared glasses without showing any concentration quenching. The increase in the intensity of emission and excitation spectra justifies the increase in Eu^{3+} ion luminescence centers with the increase in Eu^{3+} ion concentration. Among all glasses, SbPbOFBEu2.5 glass had very intense peaks compared to other glasses. The intensities of the transitions are controlled by J-mixing.[29] The $^5D_0 \rightarrow {}^7F_4$ (701 nm) transition situated in near infrared region showed relatively less intensity when compared with the $^5D_0 \rightarrow {}^7F_2$ (592 nm) transition observed in the visible red region. The $^5D_0 \rightarrow {}^7F_3$ transition was forbidden in accordance with J-O theory. However, the relatively lower intensity observed in the PL spectrum was because of the J-mixing.[29] J-O parameters Ω_λ (λ= 2, 4 and 6) are further

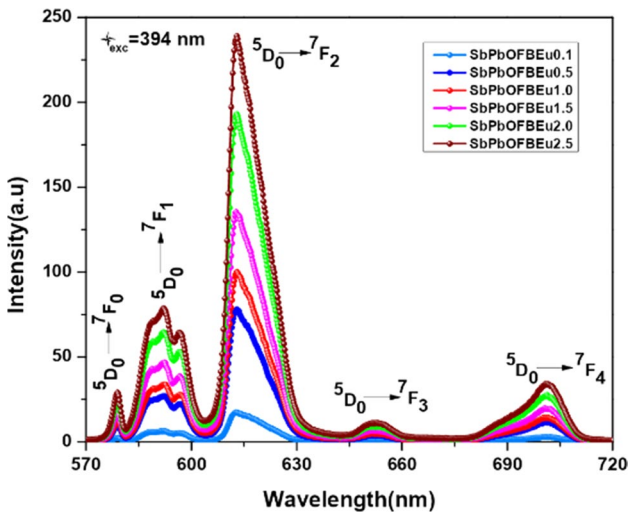


Fig. 5 Emission spectra of Eu^{3+} ion-doped SbPbOFB glasses.

useful in evaluating the emission properties for radiative transitions. From emission spectra, J-O parameters for Eu^{3+} ion-doped SbPbOFB glasses were calculated and are given in Table II along with other reported values.[1,13,30–33] The J-O parameters for all the glass samples follows the same trend ($\Omega_2 > \Omega_4$). The Ω_2 J-O parameter signifies bonding among Eu^{3+} ions along with surrounding ligands that is mostly covalent in nature. From J-O parameters, it was found that SbPbOFBEu0.1 glass had less magnitude while SbPbOFBEu2.5 glass had relatively high covalency compared to the other glasses. The J-O parameters Ω_λ (λ=2, 4 and 6) of the Eu^{3+} ions could be estimated from the ratio of intensities of $^5D_0 \rightarrow {}^7F_J$ (J=2, 4 and 6) transitions $\left[\int I_J d\vartheta\right]$ to the intensity of $^5D_0 \rightarrow {}^7F_1$ transition $\left[\int I_J d\vartheta\right]$ employing the following expression[34]:

$$\frac{\int I_J d\vartheta}{\int I_1 d\vartheta} = \frac{A_J}{A_1} = \frac{e^2}{S_{md}} \frac{\vartheta_J^3}{\vartheta_1^3} \frac{n\left(n^2+2\right)^2}{9n^3} \Omega_\lambda < \varphi_J \left|\left| \cup^\lambda \right|\right| \varphi'_{J'} >^2 \quad (1)$$

where ϑ_1 is the wavenumber (cm^{-1}) for the $^5D_0 \rightarrow {}^7F_1$ transition while ϑ_J is the wavenumber for the $^5D_0 \rightarrow {}^7F_J$ (J = 2, 4 and 6) transitions, respectively. S_{md} is the magnetic dipole line strength of the $^5D_0 \rightarrow {}^7F_1$ transition of the Eu^{3+} ion that was unique to the host glass matrix. In the present work, we have neglected the usage of Ω_6 J-O parameter values in the evaluation of radiative parameters as their impact is minor.[35–37]

PL intensities were enhanced once the bond present between the Eu^{3+} ions and oxygen were more covalent. For understanding the covalent/ionic power linking the Eu^{3+} ions with the surrounding ligands and to find the degree of the asymmetry close to local situation of the Eu^{3+} ions, the PL intensity ratio, $R = (^5D_0 \rightarrow {}^7F_2)/(^5D_0 \rightarrow {}^7F_1)$ is evaluated. The R values are evaluated from the area under the peak of $^5D_0 \rightarrow {}^7F_J$ luminescence transitions for all glasses. From Table II, it can be seen that R/O values increase with the increase in concentration of the Eu^{3+} ions, and the obtained values did not matched with other reported values in the literature. The highest R/O value is for SbPbOFBEu2.5 glass, suggesting that the Eu^{3+} ions have less symmetry and greater Eu–O covalence.

Radiative Properties Analysis

By using J-O parameters, several radiative properties such as transition probabilities (A_R), total transition probabilities (A_T), and branching ratios (β_R and β_{exp}), stimulated emission cross-sections (σ_{se}), gain band width ($\sigma_{se} \times \Delta\lambda_p$) and optical gain ($\sigma_{se} \times \tau_R$) for the prominent transitions such as $^5D_0 \rightarrow {}^7F_1$, $^5D_0 \rightarrow {}^7F_2$ and $^5D_0 \rightarrow {}^7F_4$ are evaluated using the equations given in the literature and the results are displayed in Table III.[38]

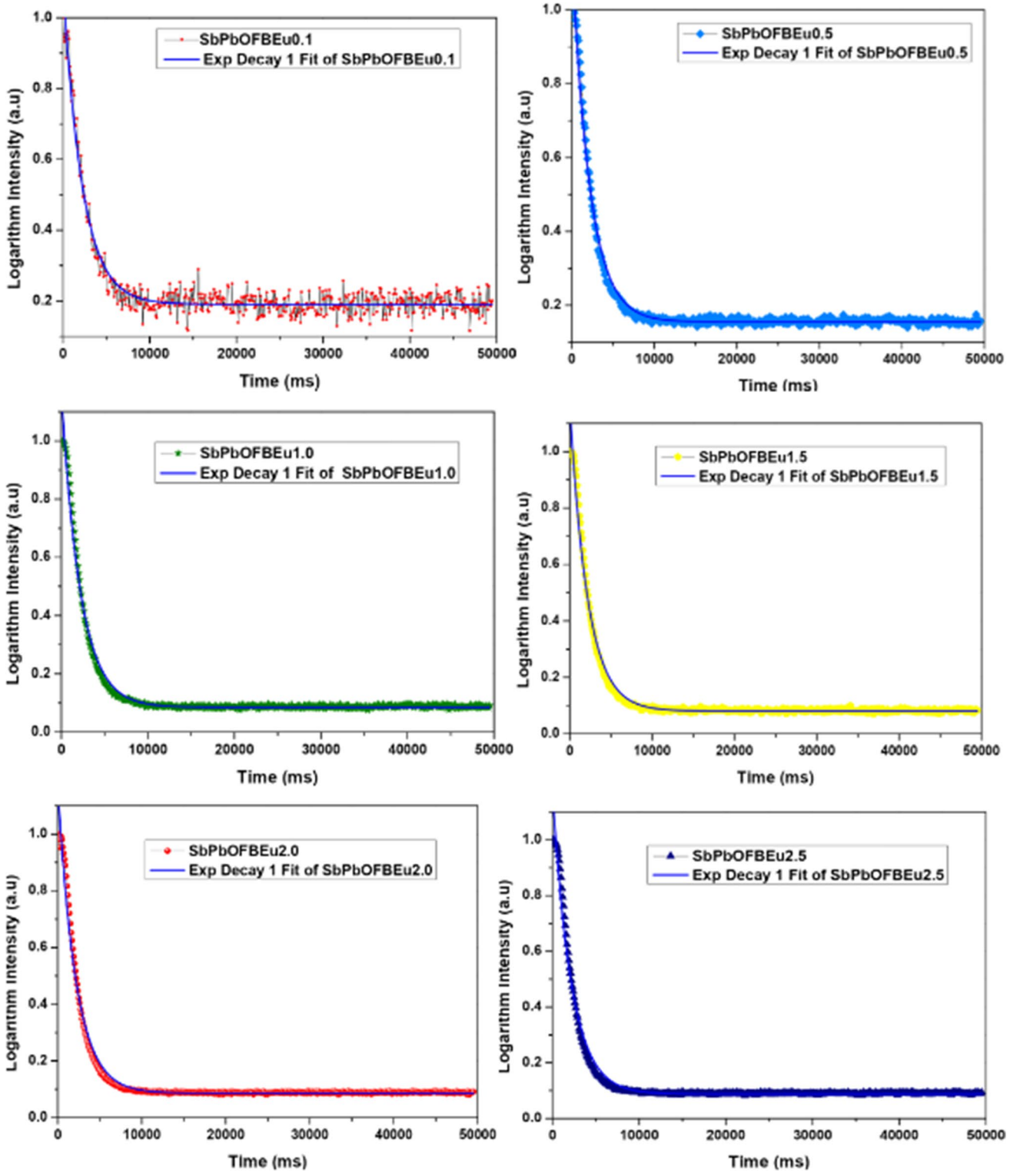


Fig. 6 PL decay spectra of Eu^{3+} ion-doped SbPbOFB glasses.

The radiative parameters are advantageous for lasing potentiality of electric-dipole transitions within the excited state to ground state. A_R could be evaluated by initial a_J and final a_J[39,40] as

$$A = \frac{64\pi^4\lambda_0^3}{3h(2J+1)}\left[\frac{n\left(n^2+2\right)^2}{9}S_{ed} + n^3 S_{md}\right] \quad (2)$$

where λ_0 is the peak wavelength of the emission band, whereas S_{ed} and S_{md} were electric dipole line strength and the magnetic dipole line strength, respectively. In total, summation of the transition probability for the higher state to its lower state results in A_T.

$$A_T = \sum A \quad (3)$$

Radiative lifetime (τ_R) for the excited state could be found by utilizing the following expression:

$$\tau_R = \frac{1}{A_T} \quad (4)$$

Radiative branching ratio (β_R) is another important parameter to recognize the optical potentiality behaviour of the as-prepared glass and can be evaluated by using the below equation:

$$\beta_R = \frac{A}{A_T} \quad (5)$$

It is well known that the best laser glass has relatively greater stimulated emission cross-section (σ_{se}) which can be evaluated by means of the following expression[41]:

$$\sigma_{se} = \frac{\lambda_p^4}{8\pi c n^2 \Delta\lambda_p} A \quad (6)$$

Table III depicts the emission peak wavelength (λp), A_R, A_T for all Eu^{3+} iions in SbPbOFB glass samples. From Table III it is seen that the values of A_R and A_T are gradually increasing for the studied glasses with an increase in Eu^{3+} ion concentration. Among all the glasses studied in the present work, SbPbOFBEu2.5 glass holds relatively higher values for A_R and A_T for $^5D_0 \rightarrow {}^7F_2$ transition. Table III also represents branching ratio (β_{exp} and β_{cal}), σ_{se}, effective bandwidth ($\Delta\lambda_p$), gain bandwidth ($\sigma_{se} \times \Delta\lambda_P$), and the optical gain ($\sigma_{se} \times \tau_R$) values calculated for all PL transitions in the titled glasses. The experimental branching ratio (β_{exp}) is determined using the area under the emission peak, and the radiative branching ratio (β_R) values are calculated by employing the J-O theory. Then, β_{exp} values were found to be very high for the $^5D_0 \rightarrow {}^7F_2$ transition in SbPbOFBEu2.5 glass in the range of 0.50. The branching ratio characterizes the lasing strength of a PL transition if its values are ≥ 0.50.[42] Also, lasing potentiality of a glassy material can be understood using the estimated σ_{se} values. From Table III it is seen that σ_{se} values are relatively higher for the $^5D_0 \rightarrow {}^7F_2$ transition of SbPbOFBEu2.5 glass than the other Eu^{3+} ion-doped SbPbOFB glasses.

PL Decay Spectral Analysis

Figure 6 illustrates the PL decay curves recorded for $^5D_0 \rightarrow {}^7F_2$ emission transition observed at 613 nm for the studied glasses. All the decay curves are well fitted to single exponential $\left(I = I_0 \exp^{(-t/t)}\right)$ and the experimental lifetime (τ_{exp}) values are shown in Table IV along with radiative lifetimes (τ_R). The luminescence quantum efficiency (η) of a material is defined as the ratio for number of photons emitted to the number of photons absorbed. The value of η is useful in understanding the lasing efficiency of the glass matrix

Table II The Judd-Ofelt intensity parameters (Ω_λ x $10^{-22}cm^2$) evaluated from the photoluminescence spectra of the Eu^{3+} ion-doped SbPbOFB glasses

Name of the glass	Ω_2 [a]	Ω_4 [b]	Ω_6	Trends	Intensity ratio (R/O)	References
SbPbOFBEu0.1	0.79±0.01	0.380±0.01	–	$\Omega_2 > \Omega_4$	2.73	Present Work
SbPbOFBEu0.5	0.82±0.01	0.404±0.01	–	$\Omega_2 > \Omega_4$	2.82	Present Work
SbPbOFBEu1.0	0.99±0.01	0.406±0.01	–	$\Omega_2 > \Omega_4$	2.88	Present Work
SbPbOFBEu1.5	1.02±0.01	0.412±0.01	–	$\Omega_2 > \Omega_4$	2.94	Present Work
SbPbOFBEu2.0	1.10±0.01	0.422±0.01	–	$\Omega_2 > \Omega_4$	3.06	Present Work
SbPbOFBEu2.5	1.11±0.01	0.434±0.01	–	$\Omega_2 > \Omega_4$	3.07	Present Work
1.50Eu	14.5	0.45	–	$\Omega_2 > \Omega_4$	–	29
Au0.15	7.55	2.42	–	$\Omega_2 > \Omega_4$	–	30
Au0.05	6.41	2.70	–	$\Omega_2 > \Omega_4$	–	30
B_2O_3-GeO_2-Gd_2O_3	5.45	4.45	–	$\Omega_2 > \Omega_4$	–	31
CaSiO3Eu0.1	3.54	0.00	–	$\Omega_2 > \Omega_4$	–	32
TeWK	2.08	0.89	–	$\Omega_2 > \Omega_4$	–	13
E3	1.98	1.41	–	$\Omega_2 > \Omega_4$	–	1

[a] Determined from the transitions $^5D_0 \rightarrow {}^7F_2$ (emission spectra).

[b] Determined from the transitions $^5D_0 \rightarrow {}^7F_4$ (emission spectra)

under investigation. Practically, the η value for the studied glasses can be estimated using the following equation.[43]

$$\eta = \frac{\tau_{exp}}{\tau_R} \times 100\% \quad (7)$$

The η values estimated for the studied glasses are given in Table IV. From Table IV, it can be observed that the η value of the glasses are gradually increasing with Eu^{3+} ion concentration.

It is observed from Table IV that the glasses possess relatively higher lifetime values compared to the other reported glasses.[1,13,44] From the estimated parameters such as β_{exp}, σ_{se} and η it was observed that SbPbOFBEu2.5 glass is the most appropriate among all the glasses investigated for giving bright red-light emission at 613 nm. Amongst all glasses, the BPbSbNaFEu2.5 glass has the maximum quantum efficiency (96.4%).

From Table IV , it can be observed that the measured lifetimes are decreasing with an increase in Eu^{3+} ion concentration in the studied glasses due to the cross-relaxation process. The Eu^{3+} ions in the higher excited level experience both radiative and non-radiative decay transitions.

Colorimetric Analysis

The evaluation and quantification of the color emitted by the material is known as colorimetry or the science of color. The blending of similar colors appear as one color to the human eye and it is impotent to identify the original bicolored configuration of the color.[45] The science of color mixing is expressed by a mathematical formula known as the color-matching function and is characterized by CIE chromaticity coordinates. The red-light emission from this set of glass samples was fixed by CIE1931color chromaticity coordinates as shown in Fig. 7. The color of luminescent sources can be identified from the color-matching function values such as $\overline{X}(\lambda)\overline{Y}(\lambda)and\overline{Z}(\lambda)$ and its degree of the stimulation

Table IV The experimental (τ_{exp}) (ms) and radiative (τ_R) (ms) lifetimes and quantum efficiencies (η) for Eu^{3+} ion-doped SbPbOFB glasses excited at a wavelength of 394 nm.

Name of the glass	τ_{exp}	τ_R	η	References
SbPbOFBEu0.1	2234	2680	83.3	Present Work
SbPbOFBEu0.5	2180	2605	83.6	Present Work
SbPbOFBEu1.0	2194	2397	91.5	Present Work
SbPbOFBEu1.5	2176	2356	92.3	Present Work
SbPbOFBEu2.0	2160	2267	95.2	Present Work
SbPbOFBEu2.5	2154	2243	96.0	Present Work
TeWLiK	3241	1075	33	13
E3	–	–	47	1
E5	–	–	55	1
NaTFP	–	–	70	44
TeWK	1737	–	83	13

Table III Wavelength (nm), transition probabilities (A_R) (s^{-1}), total transition probabilities (A_T) (s^{-1}), radiative and experimental branching ratios (β_R and β_{exp}), stimulated emission cross-sections (σ_{se}) ($\times 10^{-22}cm^2$), gain bandwidth ($\sigma_{se} \times \Delta\lambda_p$)($\times 10^{-20}cm^2$) and optical gain ($\sigma_{se} \times \tau_R$)($\times 10^{-17}$ cm^2) for $^5D_0 \rightarrow {}^7F_1$, $^5D_0 \rightarrow {}^7F_2$, $^5D_0 \rightarrow {}^7F_4$ luminescence transitions of Eu^{3+} ion-doped SbPbOFB glasses

Transition	Parameters	SbPbOFBEu0.1	SbPbOFBEu0.5	SbPbOFBEu1.0	SbPbOFBEu1.5	SbPbOFBEu2.0	SbPbOFBEu2.5
$^5D_0 \rightarrow {}^7F_1$	λ_P	592	592	592	592	592	592
	A_R	203.6	205.4	206.5	207.2	208.2	209.0
	A_T	373.0	383.7	417.1	424.3	441.0	445.6
	βR	0.546	0.535	0.495	0.488	0.472	0.469
	βexp	0.246	0.240	0.236	0.238	0.235	0.232
	σse ($\times 10^{-22}cm^2$)	4.096	4.081	4.110	4.114	4.091	4.123
	$\sigma_{se} \times \Delta\lambda_p$($\times 10^{-20}cm^2$)	0.563	0.564	0.565	0.566	0.567	0.568
	$\sigma_{se} \times \tau_R$ ($\times 10^{-17}$ cm^2)	1.244	1.206	1.171	1.155	1.129	1.121
$^5D_0 \rightarrow {}^7F_2$	λ_P	613	613	613	613	613	613
	A_R	136.6	143.0	174.8	180.8	195.4	197.9
	A_T	373.0	383.7	417.1	424.3	441.0	445.6
	β_R	0.366	0.372	0.419	0.426	0.443	0.444
	β_{exp}	0.632	0.644	0.653	0.650	0.655	0.654
	σ_{se} ($\times 10^{-22}cm^2$)	5.848	6.114	6.370	6.498	6.650	6.780
	$\sigma_{se} \times \Delta\lambda_p$($\times 10^{-20}cm^2$)	0.683	0.706	0.745	0.762	0.791	0.798
	$\sigma_{se} \times \tau_R$ ($\times 10^{-17}$ cm^2)	1.510	1.509	1.542	1.555	1.575	1.575
$^5D_0 \rightarrow {}^7F_4$	λ_P	701	701	701	701	701	701
	A_R	32.74	35.24	35.70	36.31	37.46	38.68
	A_T	373.0	383.7	417.1	424.3	441.0	445.6
	β_R	0.087	0.091	0.085	0.085	0.084	0.086
	β_{exp}	0.121	0.114	0.109	0.111	0.109	0.112
	σ_{se} ($\times 10^{-22}cm^2$)	1.370	1.560	1.517	1.554	1.592	1.639
	$\sigma_{se} \times \Delta\lambda_p$($\times 10^{-20}cm^2$)	0.184	0.211	0.214	0.217	0.220	0.224
	$\sigma_{se} \times \tau_R$ ($\times 10^{-17}$ cm^2)	0.407	0.451	0.444	0.443	0.439	0.443

Table V The obtained CIE chromaticity coordinates (x, y) and correlated color temperature of the Eu^{3+} ion-doped SbPbOFB glasses along with the values reported for other glasses

Eu^{3+} ion concentration (in SbPbOFB glasses)	Coordinates		References
	X	Y	
SbPbOFBEu0.1	0.6333	0.3646	Present Work
SbPbOFBEu0.5	0.6417	0.3573	Present Work
SbPbOFBEu1.0	0.6400	0.3588	Present Work
SbPbOFBEu1.5	0.6398	0.3590	Present Work
SbPbOFBEu2.0	0.6419	0.3571	Present Work
SbPbOFBEu2.5	0.6431	0.3561	Present Work
Au0.05	0.3976	0.3976	30
Au0.15	0.3836	0.3836	30
GEu	0.6080	0.3880	46
GEuAg40	0.6180	0.3790	46
TeWLiK	0.6400	0.3500	13

is vital for color matching for the spectral power density P(λ). These tri-stimulus values were used in calculating CIE1931 color chromaticity coordinates $\left(x = \frac{X}{X+Y+Z}, y = \frac{Y}{X+Y+Z}\right)$. The CIE chromaticity coordinates for Eu^{3+} ions in SbPbOFB glasses under 394-nm excitation wavelength were estimated using the abovementioned procedure and are given in Table V. From Fig. 7, it can be clearly understood that regardless of the Eu^{3+} concentration in the glasses, the color coordinates calculated for all the glasses fall in the bright red region. Table V shows the obtained CIE chromaticity coordinates (x,y) values for the as-prepared glass samples along with the values reported in the literature.[13,30,46]

Conclusions

SbPbOFB glasses doped with Eu^{3+} ions were prepared by utilizing rapid melt quenching and characterized using various spectroscopic techniques to understand their practical utility in visible red photonic device applications. Oscillator strengths measured from the absorption spectral features were subjected to J-O analysis to estimate the radiative properties for the observed fluorescence levels. Various excitation bands observed in the UV region validate the utility of commercially accessible UV green LEDs as an excitation source for the studied glasses. Then the PL spectrum measured for the glasses show five bands at 578, 592, 613, 652 and 701 nm pertaining to $^5D_0 \rightarrow {}^7F_J$ ($J = 0, 1, 2, 3, 4$) transitions, respectively. The red to orange luminescence intensity ratio (R/O) increases with the Eu^{3+} concentration in the as-prepared glass samples showing the decrease in symmetry of the Eu^{3+} ions while the covalence

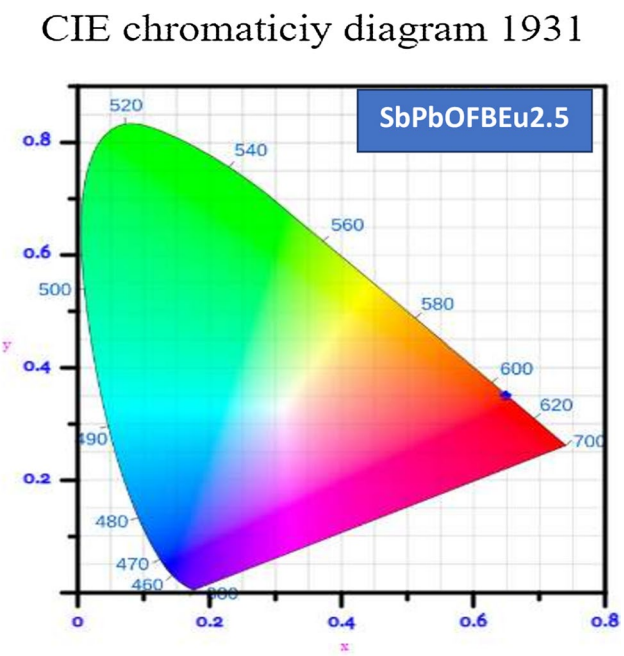


Fig. 7 CIE chromaticity diagram of 2.5 mol% of Eu^{3+} ions doped SbPbOFB glass

increases. This outcome is very reliable with the bonding property estimated from the absorption spectrum. J-O parameters resulting from the PL intensity ratio of $^5D_0 \rightarrow {}^7F_J$ ($J = 2$, 4 and 6) to the $^5D_0 \rightarrow {}^7F_2$ transitions were utilized to evaluate different radiative parameters. Amongst all the different emission transitions, the $^5D_0 \rightarrow {}^7F_2$ transition has highest J-O parameter (Ω_2) and σ_{se} representing richness of the red color in the glasses. The CIE chromaticity coordinates calculated from PL spectra reveal that the glasses with relatively higher concentration of Eu^{3+} ions are suitable for the intense bright red-light emission related to the $^5D_0 \rightarrow {}^7F_2$ transition. From the calculated σ_{se}, β_{exp},β_{cal}, η and CIE chromaticity coordinates, it is concluded that 2.5 mol% of Eu^{3+} ions doped into SbPbOFB glass is suitable for visible red photonic device applications.

Acknowledgments The authors of KLEF are very much thankful to the Department of Science and Technology (DST), Govt. of India, New Delhi, for the award of DST-FIST Level-1 (SR/FST/PS-1/2018/35) scheme to the Department of Physics, KLEF, Vijayawada, Andhra Pradesh.

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

References:

1. K. Swapna, S. Mahamuda, A. SrinivasaRao, T. Sasikala, P. Packiyaraj, L. Rama Moorthy, and G. Vijaya Prakash,

Luminescence Characterization of Eu^{3+} Doped Zinc Alumino Bismuth Borate Glasses for Visible Red Emission Applications. *J. Lumin* 156, 80–86 (2014).
2. A. Kumar Yadav and P. Singh, A Review of Structure of Oxide Glasses by Raman Spectroscopy. *RSC Adv.* 5, 67583–67609 (2015).
3. M. Fabian, F. Gergely, J. Osan, T. Cendak, S. Kesari, and R. Rao, Structural Investigation of Borosilicate Glasses Containing Lanthanide Ions. *Sci. Rep* 10, 7835 (2020).
4. A.S. Rao, J.L. Rao, and S.V.J. Lakshman, Electron Paramagnetic Resonance and Optical Absorption Spectra of Cu^{2+} ions in Alkali Cadmium Borosulphate Glasses. *J. Phy. Chem. Solids* 53, 1221–1226 (1992).
5. R. Sharma, A.S. Rao, N. Deopa, M. Venkateswarlu, M. Jayasimhadri, D. Haranath, and G. Vijaya Prakash, Spectroscopic Study of Pr^{3+} ions Doped Zinc Lead Tungsten Tellurite Glasses for Visible Photonic Device Applications. *Opti. Mat.* 78, 457–464 (2018).
6. R. Sharma, A. Prasad, S. Kaur, N. Deopa, P. Rekha Rani, M. Venkateswarlu, and A.S. Rao, Spectroscopic Properties of Deep Red Emitting Tm^{3+} Doped ZnPbWTe Glasses for Optoelectronic and Laser Applications. *J. Non- Cryst. Solids* 516, 82–88 (2019).
7. M. Nalin, M. Poula, J.L. Ribeiro, and Y. Messaddeq, Glass Formation in the Sb_2O_3-$SbPO_4$- WO_3 System. *J. Non-Cryst. Solids.* 284, 110–116 (2001).
8. P. Kostka, Z.G. Ivanova, M. Nouadji, E. Černošková, and J. Zavadil, Er-doped antimonite Sb_2O_3-PbO-ZnO/ZnS Glasses Studied by Low-Temperature Photoluminescence Spectroscopy. *J. Alloy. Compd.* 780, 866–872 (2019).
9. J. Ashok, N. Purnachand, J. Suresh Kumar, M. Srinivasa Reddy, B. Suresh, M.P.F. Graca, and N. Veeraiah, Studies on Dielectric Dispersion, Relaxation Kinetics and ac Conductivity of $Na_2OCuOSiO_2$ Glasses Mixed with Bi_2O_3-Influence of Redox Behavior of Copper Ions. *J. Alloy. Compd.* 696, 1260–1268 (2017).
10. H. Doweidar, G. El-Damrawi, and M. Abdelghany, Structure-Properties Correlations in PbF_2- B_2O_3 glasses. *Phys. Chem. Glasses Eur. J. Glass Sci. Technol. B* 55, 121–129 (2014).
11. X. Sun, and S. Huang, Nucl. Tb^{3+} -activated SiO_2–$Al2O_3$–CaO–CaF_2 Oxyfluoride Scintillating Glass Ceramics. *Instrum. Methods Phys. Res. A.* 621, 322–325 (2010).
12. D. Kumar, S.M. Rao, and S.P. Singh, Structural, Optical, and Thermos-Luminescence study of Dy^{3+} ion Doped Sodium Strontium Borate Glass. *J. Non-Cryst. Solids* 464, 51–55 (2017).
13. C.B. Annapurna Devi, S. Mahamuda, K. Swapna, M. Venkateswarlu, A. Srinivasa Rao, and G. Vijaya Prakash, Compositional Dependence of Red Luminescence from Eu3þ ions Doped Single and Mixed Alkali Fluoro Tungsten Tellurite Glasses. *Opt. Mat.* 73, 260–267 (2017).
14. Y. Wang, T. Endo, L. He, and C. Wu, Synthesis and Photoluminescence of Eu^{3+}-doped (Y, Gd) BO3 Phosphors by a Mild Hydrothermal Process. *J. Cryst. Grow.* 268, 568 (2004).
15. A.S. Rao, B.R.V. Rao, M.V.V.K.S. Prasad Kumar, J.V. Shanmukha Kumar, M. Jayasimhadri, J.L. Rao, and R.P.S. Chakradhar, Spectroscopic and Optical Properties Fluorine Containing Alkali and Alkaline Earth Zinc-Aluminophosphate Optical Glasses. *Phys. B* 404, 3717–3721 (2009).
16. S.A. Saleem, B.C. Jamalaiah, M. Jayasimhadri, A.S. Rao, K. Jang, and L.R. Moorthy, Luminescent Studies in Alkali Lead Telluro-fluoroborate Glasses. *J. Quant. Spectro. Rad. Trans.* 112, 78–84 (2011).
17. R. Akagi, N. Ohtori, and N. Umesaki, Raman Spectra of K_2O_3-B_2O_3 Glasses and Melts. *J. Non-Cryst. Solids* 293–295, 471–476 (2001).
18. D.F. Franco, and AC Sant'Ana, LFC DeOliveira, MAP Silva, The Sb_2O_3 Redox Route to Obtain Copper Nanoparticles in Glasses with Plasmonic Properties. *J. Mat. Chem.* C, 1–3 (2012).
19. A.V. Deepa, M. Priya, and S. Suresh, Influence of Samarium Oxide Ions on Structural and Optical Properties of Borate Glasses. *Sci. Res. Essays* 11, 57–63 (2016).
20. K. Knoblochova, H. Ticha, J. Schwarz, and L. Tichy, Raman Spectra and Optical Properties of Selected Bi_2O_3–PbO–B_2O_3–GeO_2 Glasses. *Opt. Mat.* 31, 895–898 (2009).
21. E.T. Yew, W.M. Hua, W.P. Sum, N.A.M. Jan, Z. Ibrahim, and R. Hussin, Structural Study of Antimony Borate Glass System Doped with Transition Metal Ions Using Infrared and Raman Spectroscopy. *J. Adv. Mater. Res.* 501, 51–55 (2012).
22. R.V. Deun, K. Binnemans, C. Gorller-Walrand, and J.L. Adam, Adam, Optical Properties of Eu^{3+} Doped Fluorophosphate Glasses. *J. Phys. Condens. Matter* 10, 7231–7241 (1998).
23. W.T. Carnall, P.R. Fields, and K. Rajnak, Electronic Energy Levels of the Trivalent Lanthanide Aquo ions IV. Eu^{3+}. *J. Chem. Phys.* 49, 4450 (1998).
24. C.R. Kesavulu, K.K. Kumar, N. Vijaya, K.-S. Lim, and C.K. Jayasankar, Thermal, Vibrational and Optical Properties of Eu^{3+} Doped Lead Fluorophosphate Glasses for Red Laser Applications. *Mater. Chem. Phys.* 141, 903–911 (2013).
25. H. Burger, K. Kneipp, H. Hobert, W. Vogel, V. Kozhukharov, and S. Neov, Glass Formation, Properties, and Structure of Glasses in the TeO_2-ZnO System. *J. Non Cryst. Solids* 151, 134–142 (1992).
26. T. Kalpana, M.G. Brik, V. Sudarsan, P. Naresh, V. Ravi Kumar, I.V. Kityk, and N. Veeraiah, Influence of Al^{3+} ions on Luminescence Efficiency of Eu^{3+} ions in Barium Boro-Phosphate Glasses. *J. Non-Cryst. Solids* 419, 75–81 (2015).
27. G. Mikhail, B. Zelika, M. Antic, K. Vukuvic, and M.D. Dramicanin, Judd-Ofelt Analysis of Eu^{3+} Emission in TiO2 Anatase. *Mater. Trans.* 56, 1416–1418 (2015).
28. W.T. Carnall, P.R. Fields, and K. Rajnak, Electronic Energy Levels in the Trivalent Lanthanide Aquo ions. I. Pr^{3+}, Nd^{3+}, Pm^{3+}, Sm^{3+}, Dy^{3+}, Ho^{3+}, Er^{3+}, and Tm^{3+}. *J. Chem. Phys.* 49, 4424–4442 (1968).
29. K. Binnemans, Interpretation of Europium (III) Spectra. *Coord. Chem. Rev.* 295, 1–45 (2015).
30. L. Yuliantini, M. Djamal, R. Hidayat, P. Yasaka, J. Kaewkhao, and K. Boonin, Development of Eu^{3+} doped Boro-Tellurite Oxyfluoride Glass and their Judd-Ofelt Analysis for Red Laser Gain Medium Application. *J. Adv. Mater. Res* 17, 1815–1822 (2019).
31. I.M. Danmallam, S.K. Ghoshal, R. Ariffin, and I. Bulus, Europium Luminescence in Silver and Gold Nanoparticles Co-embedded Phosphate Glasses: Judd-Ofelt Calculation. *Opti. Mat.* 105, 109889 (2020).
32. M. Gokçe, U. Sentürk, D.K. Uslu, G. Burgaz, Y. Sahin, and A. Gürhan Gokçe, Investigation of Europium Concentration Dependence on the Luminescent Properties of Boro Germanate Glasses. *J. Lumin.* 192, 263–268 (2017).
33. R.A. Barve, N. Suriyamurthy, B.S. Panigrahi, and B. Venkatraman, Optical Properties and Judd-Ofelt Analysis of Eu^{3+} Activated Calcium Silicate. *Phys. B Condens. Matter* 475, 156–161 (2015).
34. B. Peng, and T. Izumitani, The Fluorescence Properties of Eu^{3+} in Various Glasses and the Energy Transfer Between Eu^{3+} and Sm^{3+} in Borosilico-Phosphate Glass. *Rev. Laser Eng.* 22, 16 (1994).
35. K. Marimuthu, R.T. Karunakaran, S. Surendra Babu, G. Muralidharan, S. Arumugam, and C.K. Jayasankar, Structural and Spectroscopic Investigations on Eu^{3+} Doped Alkali Fluoroborate Glasses. *Sol. State. Sci.* 11, 1297 (2009).
36. Y. Nageno, H. Takebe, K. Morinaga, and T. Izumitani, Effect of Modifier Ions on Fluorescence and Absorption of Eu^{3+} in Alkali and Alkaline Earth Silicate Glasses. *J. Non-Cryst. Solids* 169, 288 (1994).

37. G. Tang, J. Zhu, Y. Zhu, and C. Bai, The Study on Properties of Eu^{3+} Doped Fluorogallate Glasses. *J. Alloys Compd.* 453, 487 (2008).
38. S. Mahamuda, K. Swapna, A. Srinivasa Rao, M. Jayasimhadri, T. Sasikala, K. Pavani, and L. Rama Moorthy, Spectroscopic Properties, and Luminescence Behaviour of Nd^{3+} Doped Zinc Alumino Bismuth Borate Glasses. *J. Phys. Chem. Solids* 74, 1308–1315 (2013).
39. V. Lavin, P. Babu, C.K. Jayasankar, I.R. Martin, and V.D. Rodriguez, On the local structure of Eu^{3+} ions in Oxyfluoride Glasses. Comparison with Fluoride and Oxide Glasses. *J. Chem. Phys.* 115, 10935 (2001).
40. N. Deopa, B. Kumar, M.K. Sahu, P. Rekha Rani, and A.S. Rao, Effect Sm^{3+} ionsconcentration on Borosilicate Glasses for Reddish Orange Luminescent Device Applications. *J. Non- Cryst. Solids* 513, 152–158 (2019).
41. B. Deva Prasad Raju, and C. Madhukar Reddy, Structural and Optical Investigations of Eu^{3+} in Lead Containing Alkali Fluoroborate Glasses. *Opt. Mater.* 34, 1251 (2012).
42. K. Linganna, and C.K. Jayasankar, Optical Properties of Eu^{3+} ions in Phosphate Glasses. *Spectrochim. Acta Part A* 97, 788 (2012).
43. K. Swapna, S. Mahamuda, A. Srinivasa Rao, M. Jayasimhadri, T. Sasikala, and L. Rama Moorthy, Visible Fluorescence Characteristics of Dy^{3+} Doped Zinc Alumino Bismuth Borate Glasses for Optoelectronic Devices. *Ceram. Int.* 39, 8459 (2013).
44. D. Uma Maheswari, T. Suresh Kumar, L. Rama Moorthy, K. Jang, and M. Jayasimhadri, Emission Properties of Eu^{3+} ions Ions in Alkali Telluro-Fluorophosphate Glasses. *Phys. B* 403, 1690–1694 (2008).
45. E. Fred Schubert, *Light-Emitting Diodes,2ndedition* (New York: United States of America by Cambridge University Press, 2006), pp. 292–298.
46. M. Seshadri, M.J.V. Bell, V. Anjos, and Y. Messaddeq, Influence of Silver ions in Eu^{3+} Doped Glass for Efficient Reddish-Orange and White Light Generation. *J. Alloy. Compd.* 838, 155548 (2020).

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

ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/360104742

Static analysis of multi-metal usage in grid stiffened structures

Conference Paper · April 2022

CITATIONS
0

READS
37

3 authors:



Kshitij Tripathi
Delhi Technological University
4 PUBLICATIONS **1** CITATION
SEE PROFILE



Kunal Kukreja
Delhi Technological University
3 PUBLICATIONS **1** CITATION
SEE PROFILE



Rastogi Vikas
Delhi Technological University
144 PUBLICATIONS **700** CITATIONS
SEE PROFILE

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

Project Design, development and analysis of a magnetorheological damper View project

Project Design and Development of a seven DoF Semi-Autonomous Robotic Manipulator View project

All content following this page was uploaded by Kshitij Tripathi on 22 April 2022.

The user has requested enhancement of the downloaded file.

Static analysis of multi-metal usage in grid stiffened structures

Kshitij Tripathi[1*], Kunal Kukreja[1*] and Dr Vikas Rastogi[2]
1*Student, Mechanical Engineering Department, Delhi Technological University,
2 Professor, Mechanical Engineering Department, Delhi Technological University
*Corresponding author(s). E-mail(s): kshitijtripathi18@gmail.com; kunalkukreja1002@gmail.com;
Contributing authors: rastogivikas@gmail.com;

Abstract

Structures that have components that require particularly strong strengthening without the use of large amounts of material use grid stiffened structures, modified for the purposes of the single application the structure is utilised in. Modern grid stiffened structures have evolved and deployed advanced materials, designs and manufacturing techniques. Due to the Additive Manufacturing boom in industries, any such structures manufactured with the use of metals is significantly cheaper when using superalloys for complex structures, counteracts the limitations of composite materials and is kinder on the environment.
The premise of this paper is to observe the response behaviour of such grid stiffened structures under vertical compression tests and seeks to find buckling and bending characteristics on both static and structural analyses done on ANSYS mechanical workbench. Moreover, it lays out the foundation of subsequent optimization scopes.

Keywords: Isogrid, Buckling, Metal additive manufacturing, ANSYS, Grid-Stiffened Structures

Introduction/Literature Review

In the aerospace industry, a search for the reduction of weight of rockets and other aircraft is ubiquitous. This search has led to majorly 2 methods; the first being material selection and the second, bringing changes in the design of the rocket shells. Even though the material selection process is very essential, it is limited by its metallurgical constraints and hence, can be used only to a certain extent. The focus has now been shifted to design changes. In the centre of this shift is the advent of Isogrids.
The Isogrids are characterized by: 1) Isotropic Properties; 2) High Torsional Resistance and most importantly 3) Resistance to external loading with the help of local stiffening of the pattern. [1] This structure was seen as efficient because it is well known that triangular trusses are very efficient in load distribution.

The basic cross-section of the Isogrid Geometry (with flange) was first introduced in the Isogrid Design Handbook (McDonnell-Douglas Astronautics Co., 1973).[1]

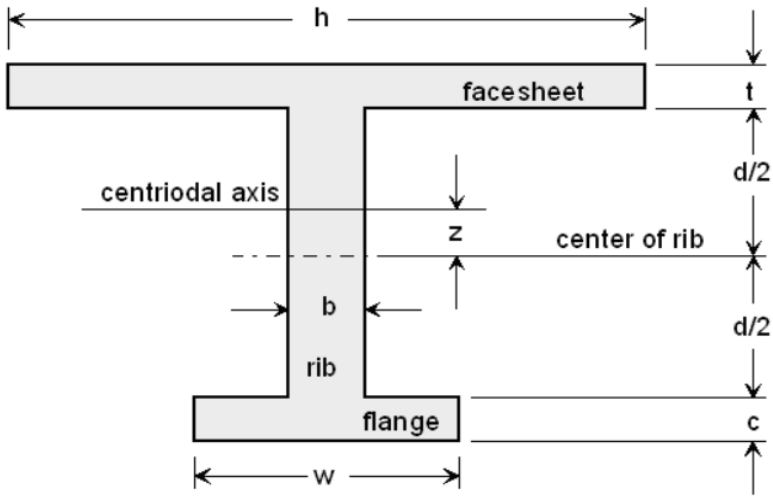


Fig 1. Cross-Section of Isogrid with Flange[1]

In Fig 1, the various dimensions of the isogrid cross-section are clearly labelled. As per the figure, h is the Cell Width or in other words, the height of the equilateral triangular cell; t is Facesheet or Skin Thickness; b and d are the Rib Width and Rib Depth respectively and w and c correspond to the Flange Width and Flange Depth. The above cross-section and dimensions prove to be helpful in the calculation of various section properties of the final isogrid plate. [1]

To manufacture such a complex geometry, a lot of expenses were used for manufacturing using conventional milling techniques. However, in recent times, additive manufacturing (or 3D printing) is a very good method of manufacturing as it is useful to manufacture complex geometries, which saves time as well as manufacturing costs. Another important observation is that additive manufacturing eliminates the requirement of assembly as it can be used to print by layered manufacturing, which omits the assembly process; again saving time and costs. (Yang J., 2015)[2].

Ti-6Al-4V was seen as a primary material for isogrids due to it having superior mechanical properties, low thermal conductivity which avoids any sort of warping and high specific properties of the lattice.[3] It excels in all sorts of performance parameters, which serve as a disadvantage to Composites as composites have low interlaminar strength and toughness, poor durability under hot-wet conditions and embrittlement due to ultraviolet (UV) rays (Boyer, R. et al) [4]. Ti-6Al-4V also exhibits good creep performance at high temperatures. Moreover, it can be printed with EBM in a high-pressure chamber while giving high tensile strength, a useful property for manufacturing. Carbon Fibre Reinforced Polymers (CFRPs) could be an alternative but they are very brittle and can easily be shattered by a hammer, which won't be suitable for isogrids (Oct 15, 2015).[5]

Another material that has been in attention is the super-alloy Inconel 718 (IN718). This material is known to have poor machinability with respect to conventional machining techniques such as conventional grinding due to its poor thermal conductivity and low elastic modulus to name a few. However, IN718 is extensively used in aerospace applications due to their high specific strength, fracture-resistant characteristics and high corrosion resistance at high temperatures (Qu, Zhang, Fang, Ye & Zhu, 2015)[6]. IN718 like any other Nickel superalloy is very useful in certain aerospace applications where excellent mechanical properties are required in elevated temperatures (Blakey-Milner et al., 2021) [7]. These properties make IN718 very suitable to be another material for isogrid manufacturing in the aerospace industry.

Methodology

An arbitrary CAD model was used to make the design of the isogrid with the following geometric dimensions:
h = 20mm, t = 1mm, d = 1mm, b = 1mm, w = 1mm, c = 1mm;

Here, h is the Cell Width or in other words, the height of the equilateral triangular cell; t is Facesheet or Skin Thickness; b and d are the Rib Width and Rib Depth respectively and w and c correspond to the Flange Width and Flange Depth.

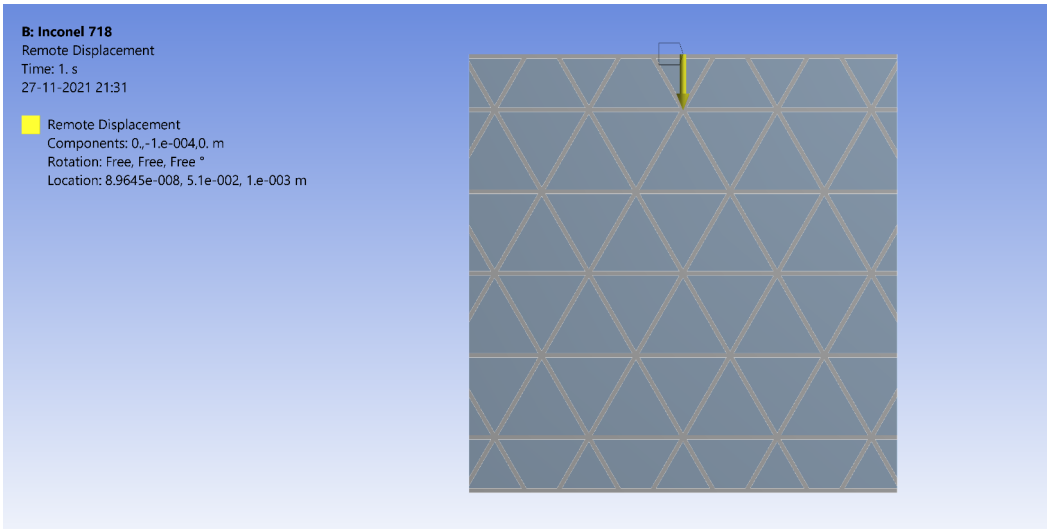


Fig 4. Remote Displacement for compression

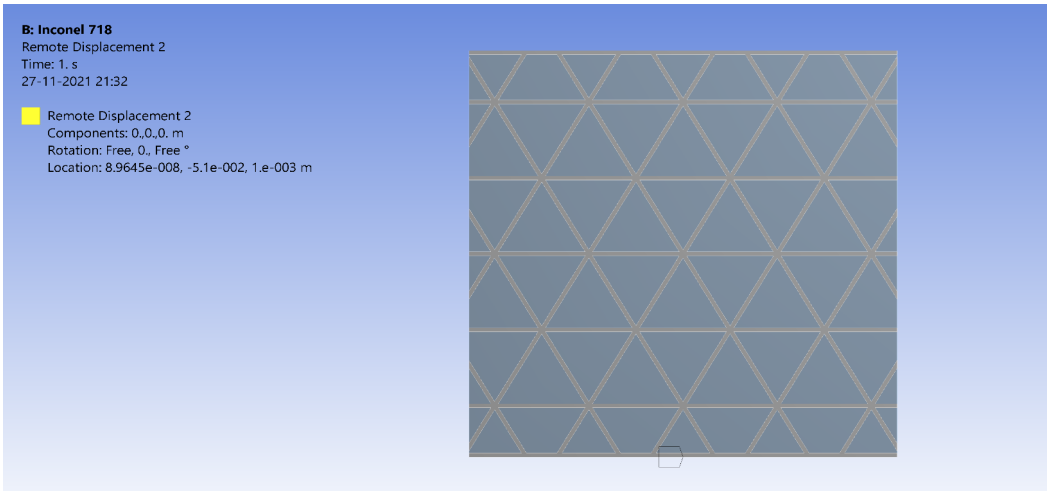


Fig 5. Remote Displacement for Support

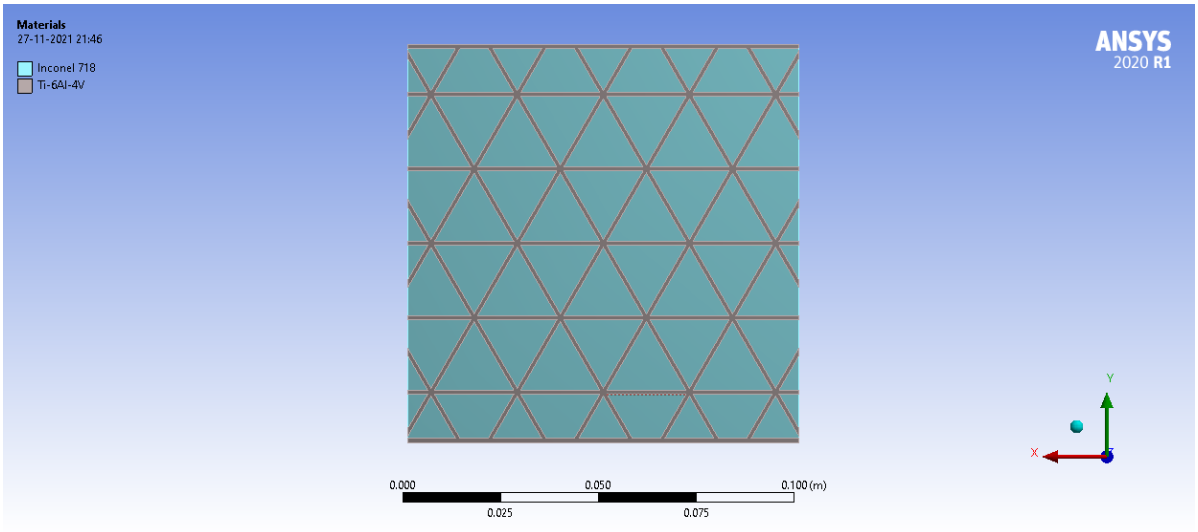


Fig 6. Material Assignment with Inconel 718 as skin and Ti-6Al-4V as ribs

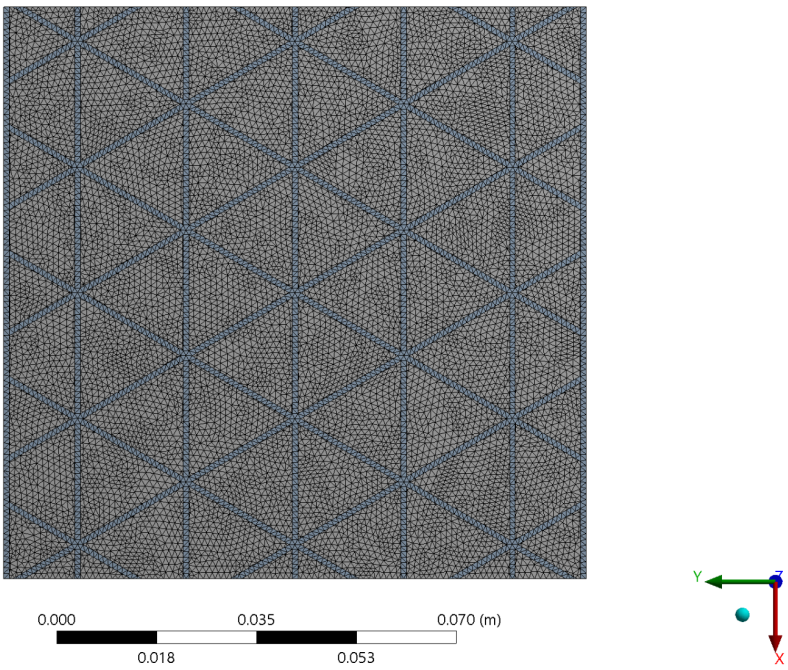


Fig 7. Refined Mesh with Tetrahedral Method of Meshing

Moreover for vertical compression tests, supporting material was added for homogenising compression surfaces to prevent end delamination of consecutive metal layers.

After doing this, based on what is desired from the isogrid to behave on the application of load, the appropriate material selection for the ribs and the skin is done. For the rib, Ti-6Al-4V is the alloy used because of its superior mechanical properties, low thermal conductivity which avoids any sort of warping and high specific properties of the lattice.

Similarly for the skin, Inconel 718 alloy was used.

For the FEA analysis on ANSYS, the structure was simply supported along with remote displacement in the skin. The meshing was in the order of 10^-9. Apart from the meshing done, an extra horizontal rib is added at both top

and bottom of the panel in order to prevent delamination. 10KPa load was applied on the top face and the equivalent stress, equivalent strain and total deformation were seen as a response to the loading.

A tetrahedral dominant mesh was preferred for the purpose of simulations of all systems, due to the fact that it is very adaptive to the geometry and all load conditions, that have and will be applied to the structure. A hex mesh is naturally less defined due to the higher volumetric efficiency at the same base cell size, increasing computation requirements, if a comparable refined hex mesh has to be obtained for a given tet mesh. Moreover, due to the subtle complexities in the structure and stress-sensitive regions, a hex mesh will be required to have prism layers near boundary conditions and the complex geometrical regions, which are more efficient if tet dominant. Hence, the conclusion was to use a tet mesh for the entire structure, making it efficient even for topology optimization.

On a comparison basis, a conventional isogrid, and a multiple-material isogrid undergo vertical compression and buckling. The resulting contours are observed and compared.

Results and Discussion

Sno.	Skin Material	Rib Material	Mesh Type	Displacement Load	Factor of Safety obtained (FOS)
1	Inconel 718	Inconel 718	Tetrahedral	10^-4m	1.1593
2	Inconel 718	Ti-6Al-4V	Tetrahedral	10^-4m	2.1147

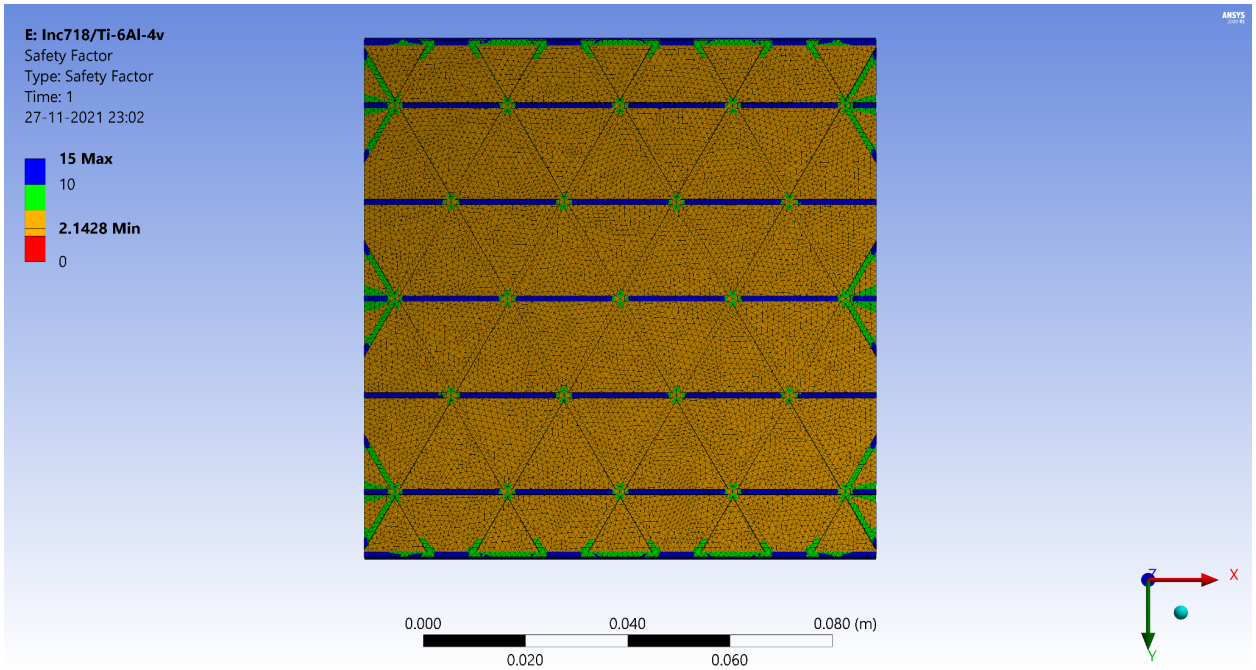


Fig 8(a)(i). Factor of Safety for Final Proposed Design (Inc/Ti)

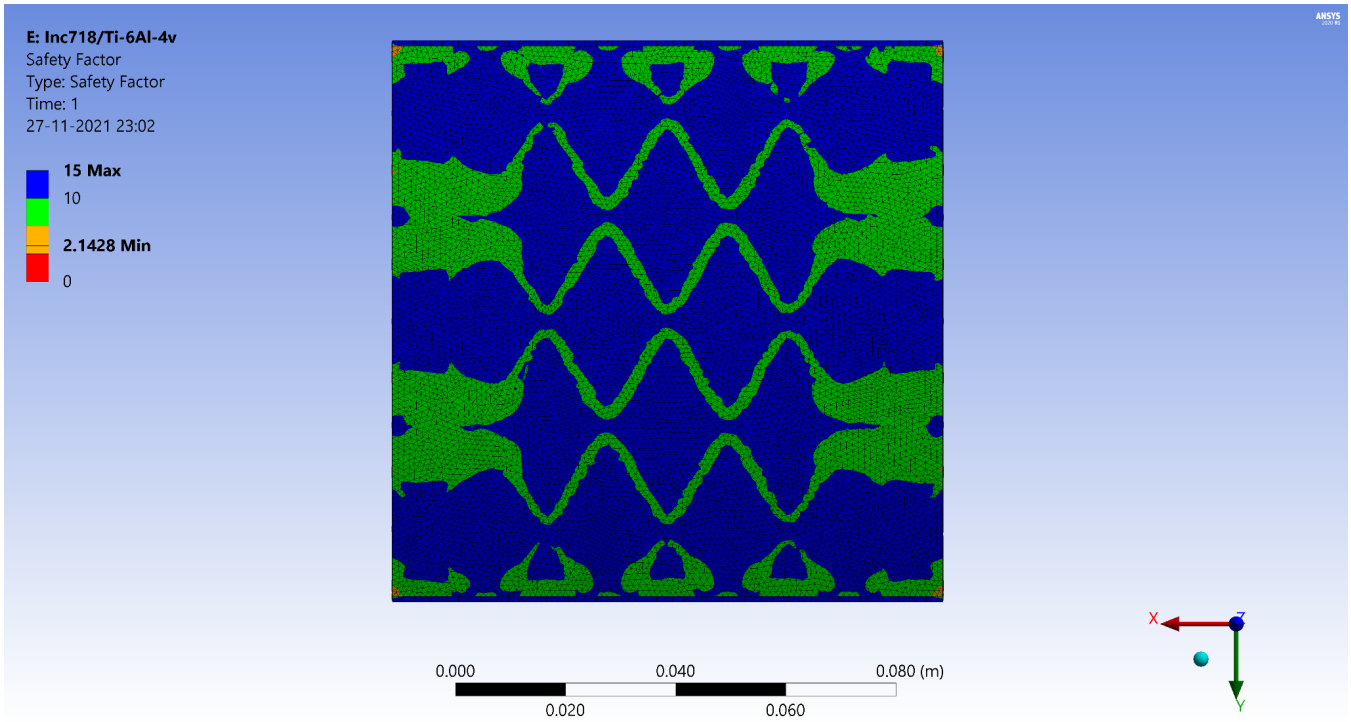


Fig 8(a)(ii). Factor of Safety for Final Proposed Design (Inc/Ti)

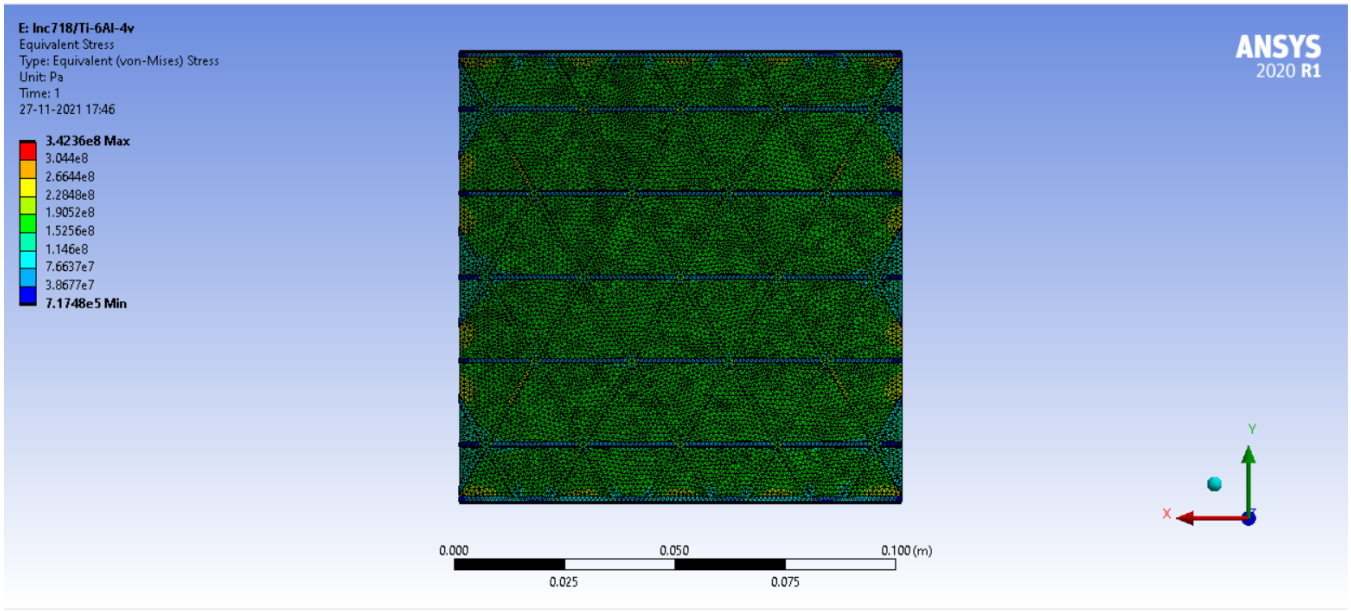


Fig 8(b). Equivalent Stress for Final Proposed Design (Inc/Ti)

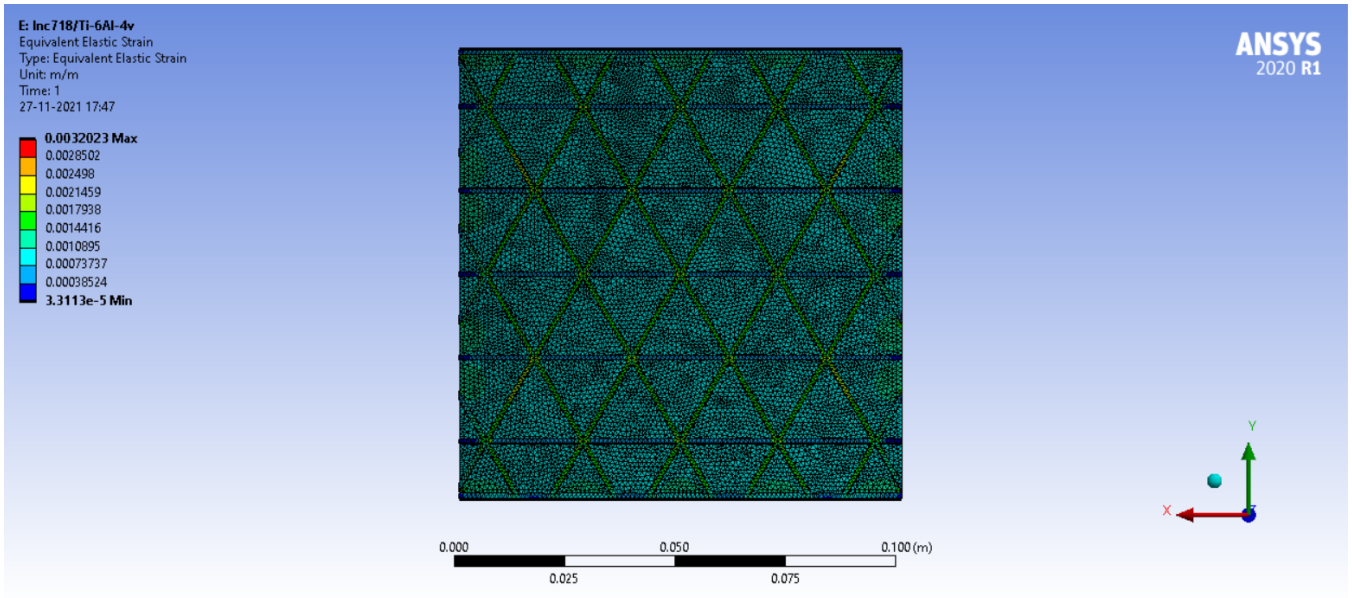


Fig 8(c). Strain for Final Proposed Design (Inc/Ti)

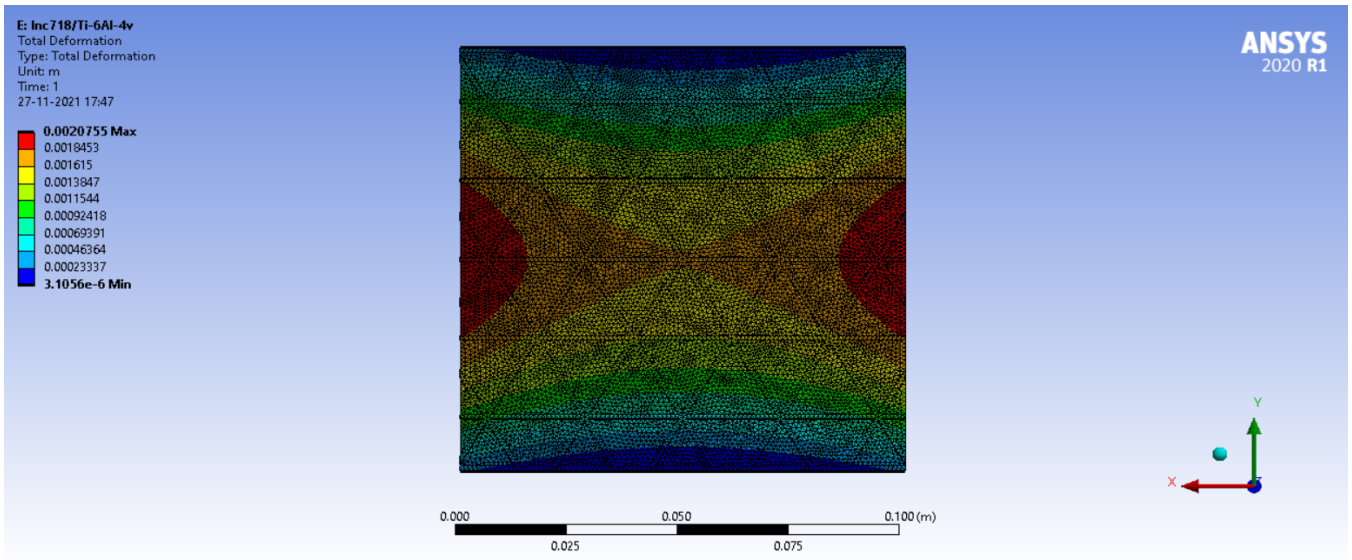


Fig 8(d). Deformation for Final Proposed Design (Inc/Ti)

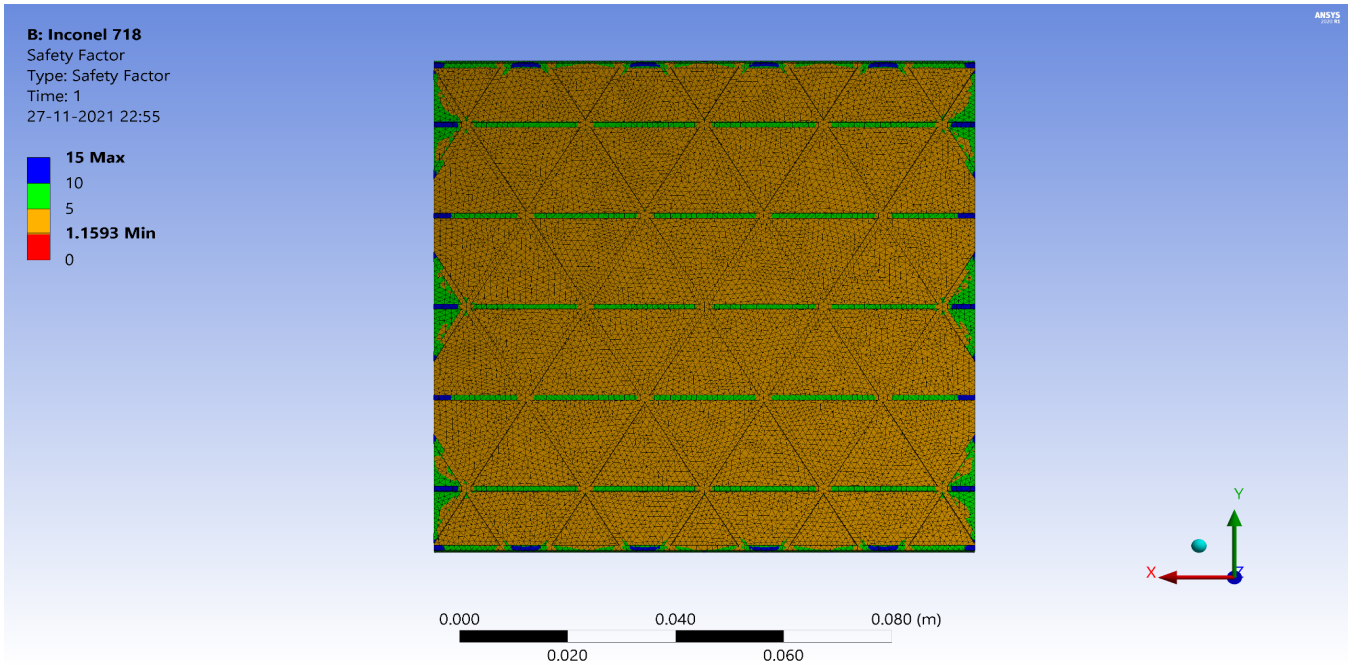


Fig 9(a)(i). Factor of Safety for Initial Proposed Design (Inc)

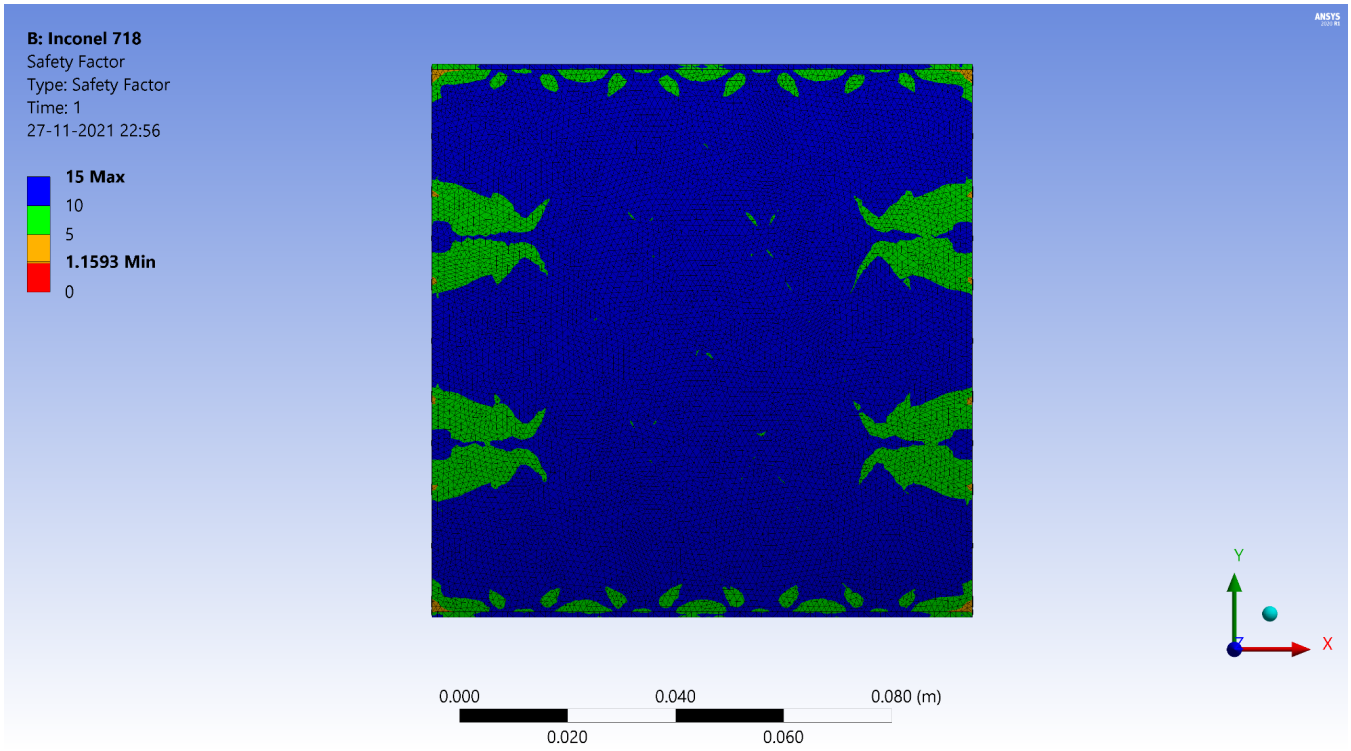


Fig 9(a)(ii). Factor of Safety for Initial Proposed Design (Inc)[back view]

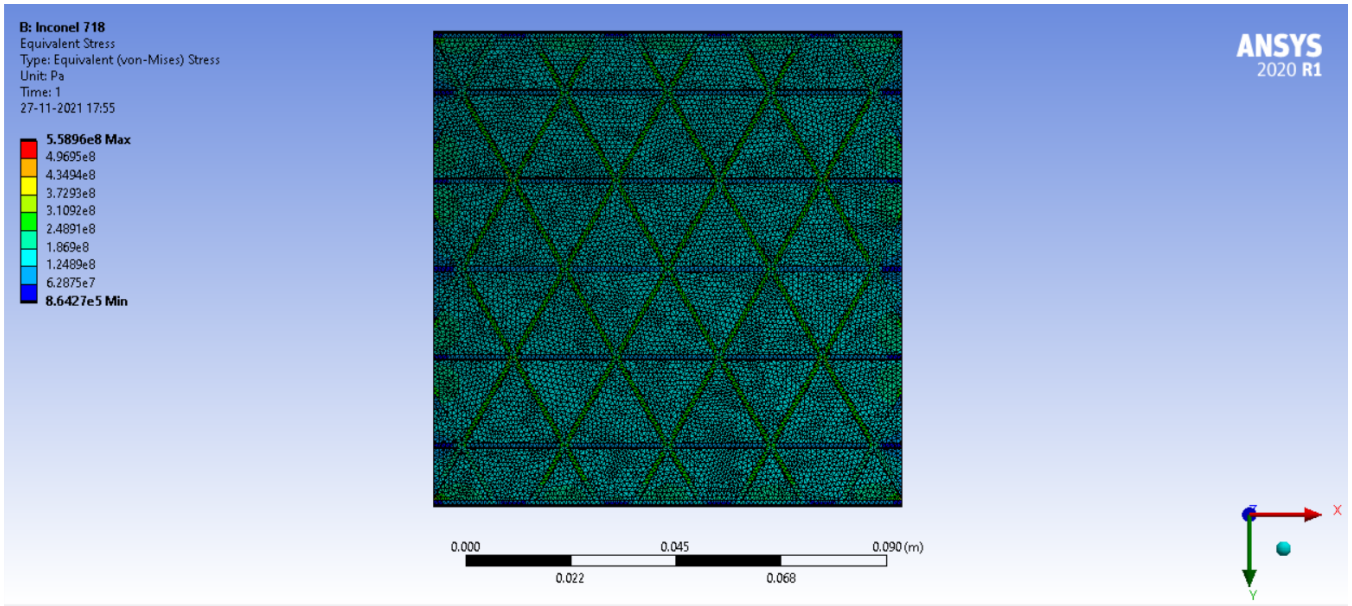


Fig 9(b). Equivalent Stress for Initial Proposed Design (Inc)

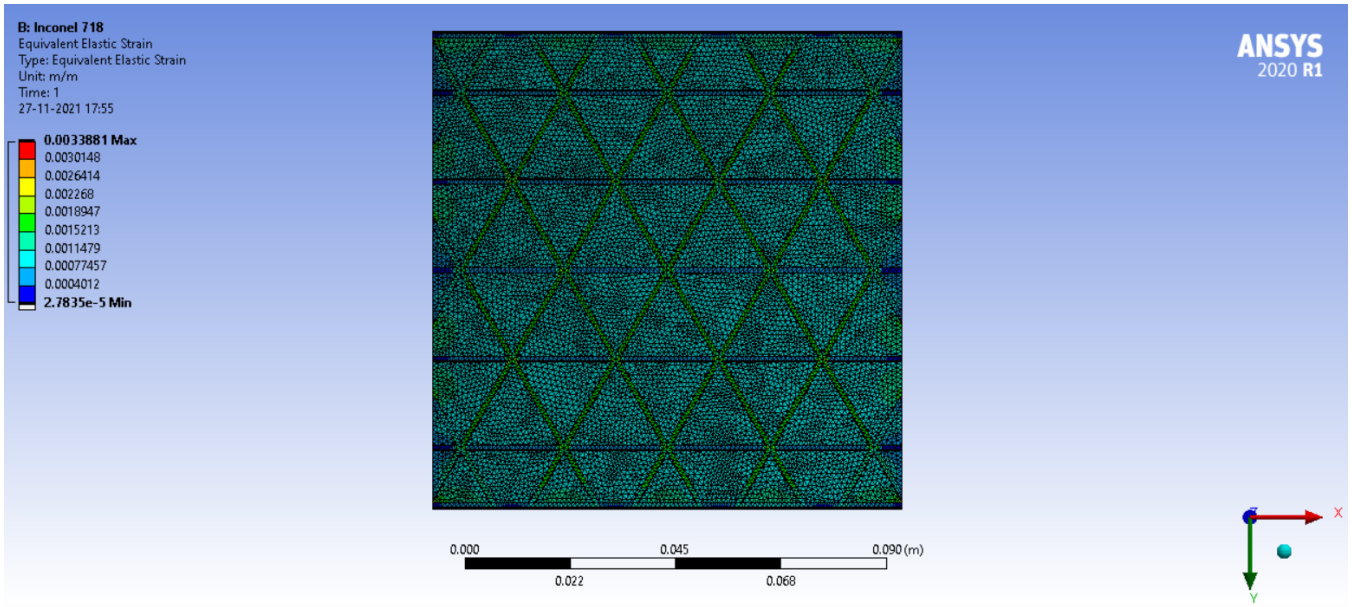


Fig9(c).Equivalent elastic strain for Initial Proposed Design (Inc)

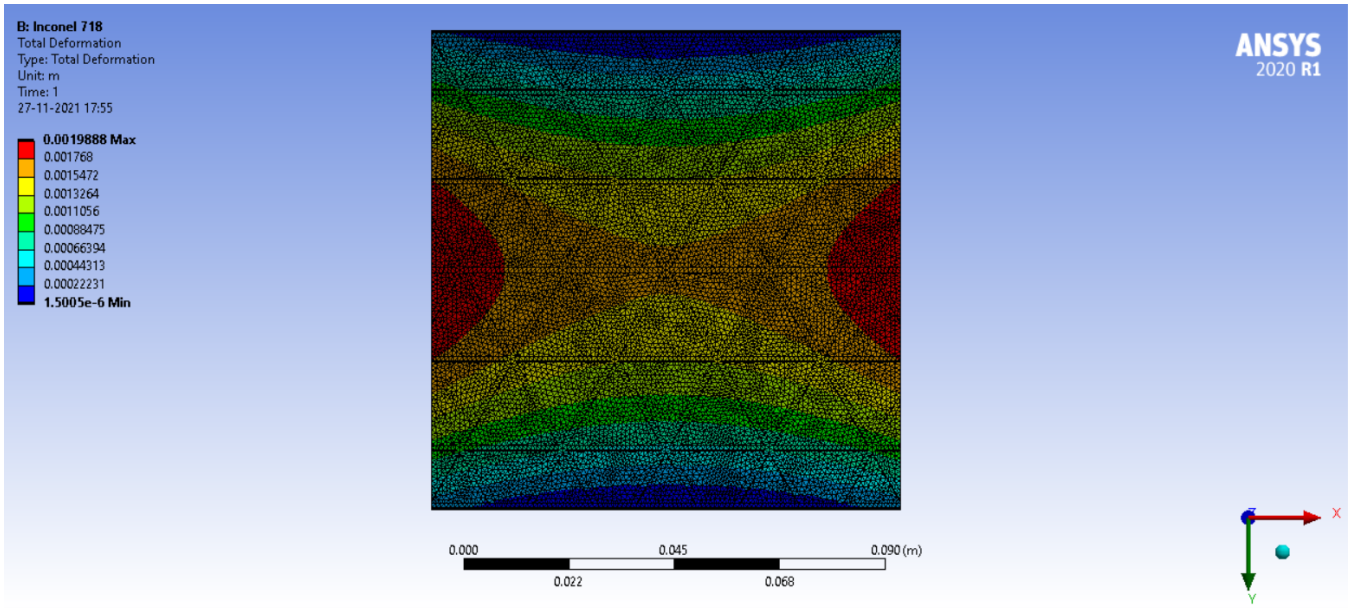


Fig 9(d). Deformation for Initial Proposed Design (Inc)

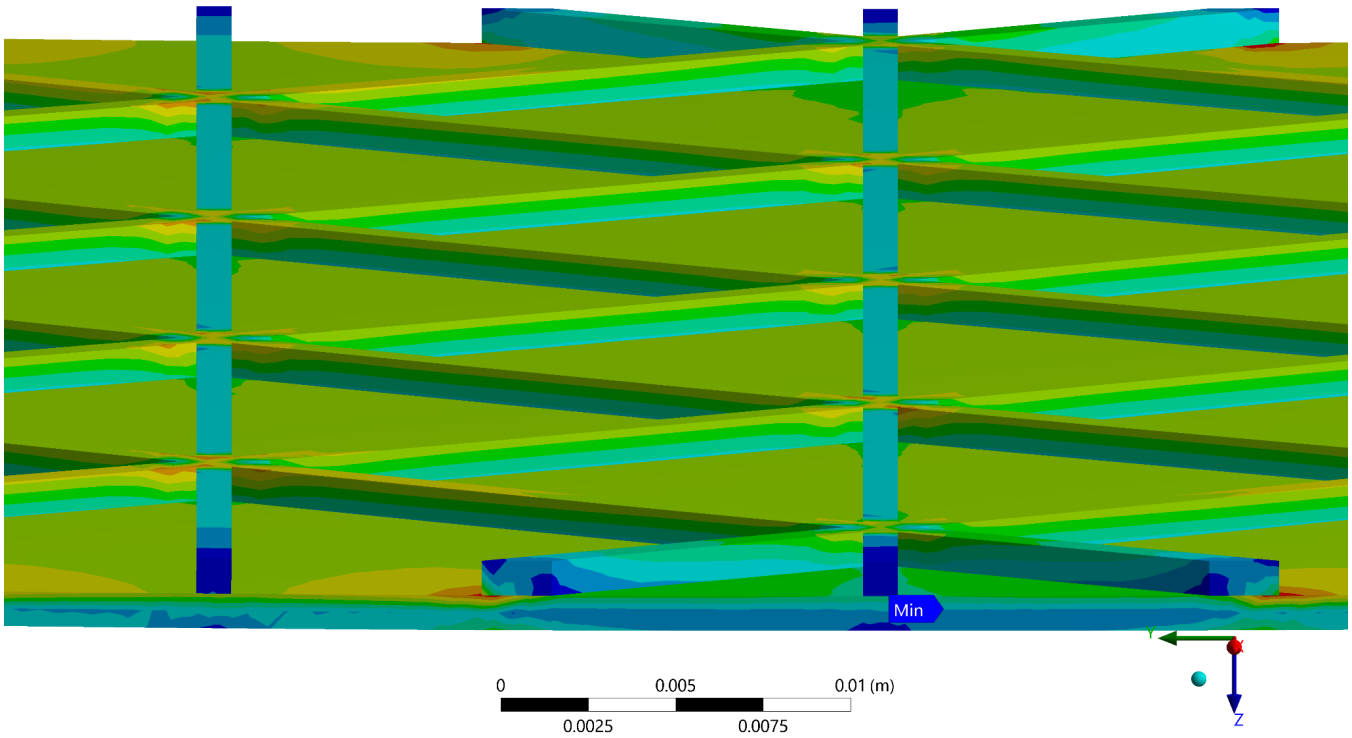


Fig 10. Arbitrary stress Stratification in Ribs and skin

A significant difference of approximately 1 with respect to the factor of safeties, in the two structures, is observed, with extremely dissimilar stress distribution characteristics. Taking this into view, case 1 and case 2 from table 1, a structure with only Inconel as a singular material was simulated (fig 8(a),(b),(c),(d)) and a proposed structure with

Ti-6Al-4V as ribs and Inconel 718 as skin (fig 9(a),(b),(c),(d)) was taken into consideration, due to the maximum disparity between Factor of safeties for the same load case.
We observe that the stress distribution in case 2 [Inc718/Ti-6Al-4V] is more effective than case 1 [Inc] as explicitly shown in Fig 8(a)(i) and Fig 9(a)(i) in the FOS contours of the simulation. We find that for case 1 the stress sustained by the nodes is significantly lesser than the rest of the ribs, portraying that they act as the focal points for stress distribution, which is not seen in case 2. Moreover, it is also evident that, due to the stress absorbed by the angled ribs, the horizontal ribs sustain very little to no stress. Finally, we see that the stress is also distributed towards the skin of the structure in case 2 Fig 8(a)(ii), at the cost of minimal deformation, whereas it is clear in case 1 Fig 9(a)(ii) that the ribs are the only load-bearing layer.
It is also worth noticing that due to the complexity of the geometry, it does not necessarily mean that the critical load before yielding, in this case, critical point taken as the yield strengths of respective material, would be approximately the factor of safety multiplied by the load applied. It does imply though, that for an identical geometric stress distribution the load for yielding would be the multiplied load. But, as we can see the load induces a stratification of stress onto the existing structure and adapts as the load varies. Hence, even though we can conclusively point out that the proposed structure gives better behaviour, the margin of better performance is subject to variable static load, and even more so with time-dependent inertial loads as shown in Fig 10.

Conclusion and future scope

The work done in this paper fundamentally lays out the premise of using purpose-specific metals in grid stiffened structures, with the help of understanding structure behaviour in ANSYS through mechanical simulations.
It is observed that on the application of a variable load, irrespective of inertial loading that a composite structure, on the usage of purpose-specific metals does give better deformation response than a conventional isogrid structure. The work also foreshadows the idea of using the same metal in multiple layers to reduce possible stress concentrations on the skin due to rib buckling and crippling and to reinforce weak rib regions while also reducing the computational time by further minimizing the use of the costlier materials by using them in quintessential areas and optimizing most of the structural complexities by using a single material. It also increases the premise of taking advantage of metal properties like metals that have higher compressive strengths compared to their tensile compressive strengths.
Further optimization studies could help minimize the use of both materials on bespoke components like brake callipers and suspension uprights used in motorsport vehicles, high sensitivity sensors and thermocouples and high-performance sports goods like cycle frames and golf clubs.
A plausible area of further development on this idea would be to integrate 3D printed surface ceramic coatings, that are in common use in the industry for a variety of resistance to external factors but are not used extensively for structural loading. Metals that are light, tough and resistive to such factors can be used in conjunction with other metals for structures that may have disadvantages for being heavy. Moreover, by printing in a non-corrosive medium and using a protective metal layer, such structures can also remove the need for auxiliary processes like anodizing which increase production time.

References

[1]]McDonnell-Douglas Aeronautics Co. Isogrid design handbook. Huntington Beach, CA, 1973

[2] Yang, J. (2015). Design and characterization of an innovative integrally stiffened structure using Additive Manufacturing (Post Graduate). Coventry University.

[3] Design of Isogrid Shells for Venus Surface Probes / Ossola, Enrico; Borgonia, John Paul; Hendry, Morgan; Sunada, Eric; Brusa, Eugenio; Sesana, Raffaella. - In: JOURNAL OF SPACECRAFT AND ROCKETS. - ISSN 0022-4650. - ELETTRONICO. - (2020), pp. 1-10. [10.2514/1.A34823]

[4] Boyer, R., Cotton, J., Mohaghegh, M., & Schafrik, R. (2015). Materials considerations for aerospace applications. MRS Bulletin, 40(12), 1055-1066. doi:10.1557/mrs.2015.278

[5] Advantages and Disadvantages of Carbon Fibre Reinforcement Polymer. (2016, Oct 15). Retrieved from ttps://studymoose.com/advantages-and-disadvantages-of-carbon-fibre-reinforcement-polymer-essay

[6] Qu, N., Zhang, Q., Fang, X., Ye, E., & Zhu, D. (2015). Experimental Investigation on Electrochemical Grinding of Inconel 718. Procedia CIRP, 35, 1. doi: 10.1016/j.procir.2015.08.055

[7] Blakey-Milner, B., Gradl, P., Snedden, G., Brooks, M., Pitot, J., & Lopez, E. et al. (2021). Metal additive manufacturing in aerospace: A review. Materials & Design, 209. doi: 10.1016/j.matdes.2021.110008

[8] Bellini, Costanzo & Borrelli, Rosario & Di Cocco, Vittorio & Franchitti, Stefania & Iacoviello, Francesco & Sorrentino, Luca. (2021). Damage analysis of Ti6Al4V lattice structures manufactured by EBM process subjected to bending load. Material Design & Processing Communications. 3. 10.1002/mdp2.223.

[9] Forcellese, A., Pompeo, V. di, Simoncini, M., & Vita, A. (2020). Manufacturing of Isogrid Composite Structures by 3D Printing. Procedia Manufacturing, 47, 1096–1100. doi:10.1016/j.promfg.2020.04.123

[10] Kim, T. D. (2000). *Fabrication and testing of thin composite isogrid stiffened panel. Composite Structures, 49(1), 21–25.* doi:10.1016/s0263-8223(99)00122-1

[11] Sorrentino, L., Marchetti, M., Bellini, C., Delfini, A., & Del Sette, F. (2017). Manufacture of high-performance isogrid structure by Robotic Filament Winding. Composite Structures, 164, 43–50. doi:10.1016/j.compstruct.2016.12.061

[12] Huybrechts, S., & Tsai, S. W. (1996). *Analysis and behaviour of grid structures. Composites Science and Technology, 56(9), 1001–1015.* doi:10.1016/0266-3538(96)00063-2

[13] Li, M., Lai, C., Zheng, Q., Han, B., Wu, H., & Fan, H. (2019). *Design and mechanical properties of hierarchical isogrid structures validated by 3D printing technique. Materials & Design, 107664.* doi:10.1016/j.matdes.2019.107664

[14] Murthy, V. C. A. D., & Santhanakrishnanan, S. (2020). *Isogrid lattice structure for armouring applications. Procedia Manufacturing, 48, e1–e11.* doi:10.1016/j.promfg.2020.05.099

[15] *Rocket Science 101: Lightweight rocket shells – Aerospace Engineering Blog*. Aerospace Engineering Blog. (2016). Retrieved from https://aerospaceengineeringblog.com/rocket-science-101-lightweight-rocket-shells/.

[16] Tripathi K., Kukreja K., Gupta N. (2022) Fabrication of Isogrids by Conventional and Unconventional Techniques: A Comparative Review Study. In: Agrawal R., Jain J.K., Yadav V.S., Manupati V.K., Varela L. (eds) Recent Advances in Industrial Production. Lecture Notes in Mechanical Engineering. Springer, Singapore. https://doi.org/10.1007/978-981-16-5281-3_41

View publication stats

[17] Wong, Kaufui. (2012). K.V. Wong, A.Hernandez, “A Review of Additive Manufacturing,” ISRN Mechanical Engineering, Vol 2012 (2012), Article ID 208760, 10 pages... ISRN Mechanical Engineering. 2012. 10.5402/2012/208760.

Materials Today: Proceedings xxx (xxxx) xxx

Contents lists available at ScienceDirect

Materials Today: Proceedings

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

Structural and spectroscopic studies of Eu^{3+} activated potassium bismuth molybdate phosphor for optoelectronic device applications

Divyanshi Nagpal, Tanisha Bhadauria, M. Jayasimhadri *

Luminescent Materials Research Lab, Department of Applied Physics, Delhi Technological University, Bawana Road, Delhi 110 042, India

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Molybdate phosphor
Morphology
Band gap
Photoluminescence

ABSTRACT

Red emitting Eu^{3+} activated $K_5BiMo_4O_{16}$ (KBM: Eu^{3+}) phosphor has been synthesized via high temperature solid state reaction route and explored via investigating various structural and spectroscopic properties. The phase purity of the as-synthesized KBM samples have been analysed through X-ray diffraction technique. Morphological properties have been analysed through SEM images and the vibration modes have been identified in KBM crystal using Fourier Transform Infrared (FT-IR) spectroscopy. The luminescence spectrum of KBM: Eu^{3+} phosphor indicates the characteristic peaks of Eu^{3+} positioned at 578, 588, 612, 657 and 704 nm under 392 and 464 nm excitation wavelength. The color coordinates of KBM: Eu^{3+} phosphor excited with 392 and 464 nm wavelength are (0.657, 0.342) and (0.655, 0.344) located in red region. Based on the above-mentioned characteristics, KBM: Eu^{3+} phosphor may be a potential candidate to utilize in optoelectronic applications such wLEDs and solar cells.
Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

1. Introduction

Among all conventional lighting sources, phosphor converted white LEDs (pc-wLEDs) render great advantages as they possess remarkable energy efficiency, good thermal stability, compact in size, high brightness, long operating lifetime, higher colour rendering index (CRI) and eco-friendliness etc [1–4]. However, pc-wLEDs also have certain issues such as low efficiency, halo effect, low color rendering index (CRI), and high color correlated temperature (CCT) value owing to deficient of red component in the crystal matrix to produce efficient white light [5]. Henceforth, the researchers and scientists are oriented to develop promising red emitting phosphor to achieve intense white light based on pc-wLEDs [6,7]. Recently, pc- wLEDs have attracted much interest in which red/green/blue phosphors excited with UV or near UV (n-UV) radiation emitting LED chip to produce white light [8].

Due to distinctive features of f-f intra configurational transitions, a trivalent rare earth (RE) ion such as Eu^{3+} doped appropriate host lattice can be used as a promising red emitting phosphor [9]. Europium ion doped phosphor has been acquiring interest due to its characteristic transitions, to employ in various applications such as display panels, lasers, solar cell and several optoelectronic devices. Moreover, appropriate host and dopant ions are required to attain qualitative luminescence. Various phosphor materials such as silicates, borates, aluminates, molybdates, vandates etc., have been explored due to the fact that they possess large band gap, high chemical and thermal stability. In this regard, alkali based bismuth molybdate hosts are presently being researched as a potential material for luminescent ions [10,11]. Since, the ionic radius of Bi^{3+} is near to the ionic radii of lanthanides, so that the bismuth containing compound can be easily doped with luminous rare earth ions [12].

In the present study, Eu^{3+} doped $K_5BiMo_4O_{16}$ (KBM) molybdate phosphor has been synthesized via solid state reaction route and explored its capability as luminescent material for optoelectronic device applications. In this direction, 1.0 mol% Eu^{3+} doped KBM phosphor has been investigated for its structural, morphological and luminescence behaviour in detail. The findings achieved in the current study endorses that the obtained red emitting KBM: $1.0Eu^{3+}$ phosphor could be an excellent candidate for w-LEDs and solar cell applications.

* Corresponding author.
E-mail address: jayasimha@dtu.ac.in (M. Jayasimhadri).

https://doi.org/10.1016/j.matpr.2022.04.432
2214-7853/Copyright © 2022 Elsevier Ltd. All rights reserved.
Selection and peer-review under responsibility of the scientific committee of the International Conference on Materials, Processing & Characterization.

Please cite this article as: D. Nagpal, T. Bhadauria and M. Jayasimhadri, Structural and spectroscopic studies of Eu^{3+} activated potassium bismuth molybdate phosphor for optoelectronic device applications, Materials Today: Proceedings, https://doi.org/10.1016/j.matpr.2022.04.432

2. Experimental section

2.1. Sample synthesis

The KBM: Eu^{3+} phosphor was synthesized using high temperature solid-state reaction route. Precursors such as potassium carbonate (K_2CO_3), bismuth oxide (Bi_2O_3), molybdenum trioxide (MoO_3) and europium oxide (Eu_2O_3) are used in stoichiometric ratio with high purity (99.8–99.9%). The precursors were dispersed in ethanol and ground for 3 h in an agate mortar and pestle. The ground sample was placed in an alumina crucible and kept in a muffle furnace for 4 h at different temperatures ranging from 800 to 1000 °C. The sintered sample was then allowed to cool down gradually till room temperature and further ground for short span. Moreover, the as-prepared phosphor powder was further characterized using characterization techniques such as XRD, DRS, SEM, FT-IR, and PL.

2.2. Characterization techniques

The XRD analysis has been done through Bruker D8 Advance model x-ray diffractometer with incident wavelength of 1.54 Å (Cu-Kα radiation). The diffraction angle recorded is associated with 40 kV of accelerating voltage and 40 mA of current in the range of 10–60 degree with an appropriate scanning rate of 0.04 s^{-1}. The band gap studies are achieved with the aid of JASCO spectrophotometer (Model no. V-770). The modes of bonding in a crystal lattice have been investigated in the range of 400 to 1600 cm^{-1} using the Perkin Elmer Spectrum Two FT-IR spectrometer. The morphological studies were performed using High Resolution Field Emission Scanning Electron Microscope (HR-FE-SEM) 7610F Plus-JEOL. Photoluminescence studies have been examined through JASCO FP-8300 spectrofluorometer established with an excitation source (Xenon lamp) for capturing both excitation and emission spectra.

3. Results and discussion

3.1. X-ray diffraction (XRD) studies

The XRD patterns of as-synthesized $K_5BiMo_4O_{16}$ (KBM) sample at different calcination temperatures ranging from 800 to 1000 °C have been represented in Fig. 1(a). The sample was calcined initially at 800 °C at which pure phase has been obtained and diffraction peaks are well matched with the standard JCPDS (card no. 29-0986) pattern. On increasing the temperature further, few impurity peaks appeared, which indicates that 800 °C is the optimal temperature for obtaining pure phase and that the sample has been crystallized well at this temperature [13,14]. The as-synthesized $K_5BiMo_4O_{16}$ hexagonal crystal has unit cell parameters a = b = 6.024 Å and c = 20.830 Å and V = 654.62 $Å^3$ [15]. Absence of any additional peak proves that no structural changes occur in forming KBM crystal. Fig. 1(b) represents the XRD pattern of 1.0 mol% of Eu^{3+} doped KBM phosphors calcined at 800 °C. The observed patterns match well with the JCPDS file as mentioned above for Fig. 1(b). The crystallite size for the as-synthesised KBM samples have been determined using the Debye-Scherrer equation [16]:

$$D = \frac{K\lambda}{\beta cos\theta} \quad (1)$$

where K is the Debye-Scherrer constant, λ is the incident wavelength of the X-ray, β is the full-width at half maxima (FWHM) and Θ is diffraction angle in radians. The crystallite sizes for undoped and 1.0 mol% of Eu^{3+} doped KBM phosphor were found to be 0.44 and 0.45 μm, respectively.

3.2. Diffuse reflectance spectroscopy (DRS)

Fig. 2 shows the optical band gap of undoped and Eu^{3+} doped KBM phosphor and have been determined using diffuse reflectance spectra in a range of 200–600 nm wavelength at RT as shown in the inset of Fig. 2. The diffuse reflectance was transformed into Kubelka Munk (K-M) function provided to estimate the optical band gap as expressed in equation below [17]:

$$F(R) = \frac{(1-R)^2}{2R} = \frac{K}{S} \quad (2)$$

where R is the ratio of sample reflectance to standard sample, and K and S denote the absorption and scattering coefficients, respectively. The linear absorption coefficient (α) in the Tauc equation has a relation with band gap energy (E_g), as shown below [18]:

$$\alpha h\upsilon = A(h\upsilon - E_g)^n \quad (3)$$

where A is proportionality constant, $h\upsilon$ represents energy of light and n is the electronic transition with value 1/2, 2, 3/2 and 3, which expresses allowed direct, allowed indirect, forbidden direct and forbidden indirect band gap with respective values of electronic transition. By comparing Eqs. (2) and (3), Tauc equation can be modified and given the expression below as:

$$(F(R)h\upsilon) = B(h\upsilon - E_g)^n \quad (4)$$

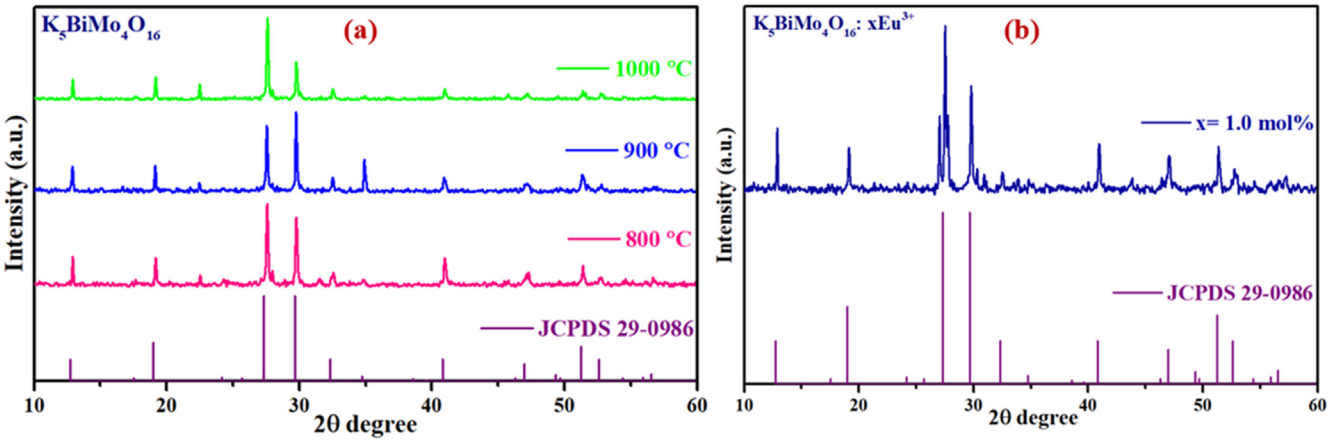


Fig. 1. (a) XRD patterns of $K_5BiMo_4O_{16}$ (KBM) samples at different calcination temperatures (800, 900 and 1000 °C for 4 h) compared with standard JCPDS data; (b) XRD pattern of 1.0 mol% of Eu^{3+} doped KBM phosphor compared with standard data.

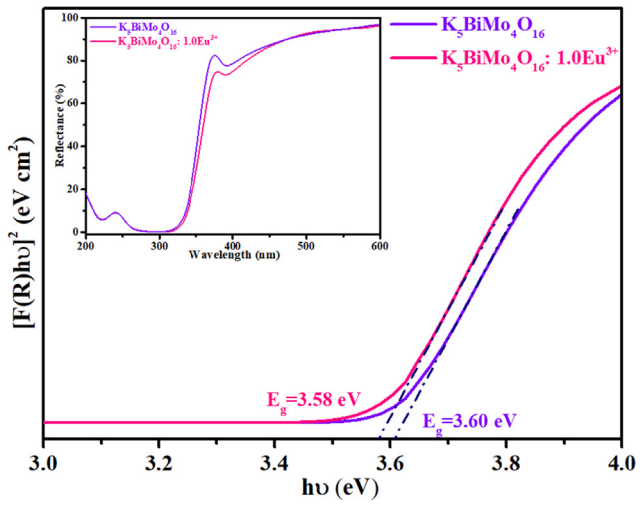


Fig. 2. Band gap spectra of undoped and 1.0 mol% of Eu^{3+} doped KBM phosphor using Kubelka-Munk function (inset: Diffuse reflectance spectra).

The energy ($h\upsilon$) vesus $(F(R)h\upsilon)^2$ plot of KBM and 1 mol% Eu^{3+} doped KBM phosphor as shown in Fig. 2, and evaluated the band gap by extrapolating the linear fit line to x-axis [19]. Using Eq. (4), the plotted energy ($h\upsilon$) vesus $(F(R)h\upsilon)^2$ graph depicts the direct allowed band gap of undoped and 1.0 mol% of Eu^{3+}doped KBM samples with the value 3.60 and 3.58 eV, respectively.

3.3. Fourier transform infrared (FT-IR) spectroscopy

The FT-IR spectrum of undoped KBM sample and 1.0 mol% Eu^{3+} doped KBM phosphor has been displayed in a range of 400–1600 cm^{-1} in Fig. 3. Both undoped and Eu^{3+} doped KBM phosphor has shown the bands at similar position with nominal shift. The characteristic bands located at specific wavenumber confirms the presence of ascribed molybdate and bismuth groups. The band at 915 cm^{-1} is due to symmetric stretching vibrations while at 852 cm^{-1} is originated from asymmetric vibration of non-bridging oxygen of Mo-O_4 and Mo-O, respectively. In FT-IR spectra, the peaks obtained in the range of 400–700 cm^{-1} is mainly due to involvement of vibration in Bi-O in the crystal lattice. The peak positioned at 689 and 453 cm^{-1} may occur owing to symmetric scissor and stretching vibrations of Bi-O_2 and Bi-O_3 group, respectively. The rocking vibration of Bi-O in crystal may reflects the presence of peaks at 516 and 565 cm^{-1} [20].

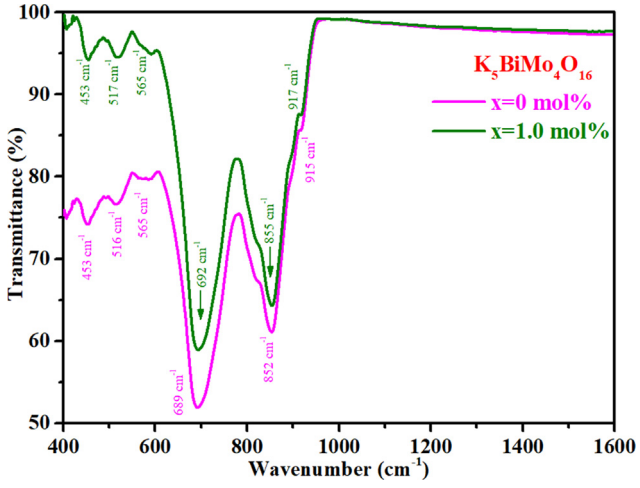


Fig. 3. FT-IR patterns of undoped and 1.0 mol% of Eu^{3+} doped KBM phosphor.

3.4. Morphological studies

The shape, morphology and particle size of sample synthesized via various physical and chemical routes can be analysed through SEM micrographs [21]. The as-prepared undoped and 1.0 mol% Eu^{3+} doped KBM phosphors at the same resolution have been shown in Fig. 4(a) and (b). The particles are in micro range and the morphology of particles in undoped and 1.0 mol% Eu^{3+} doped KBM phosphor are similar which signifies that incoorporation of dopant ion has no influence on morphology of sample [22]. The inset of Fig. 4(a) and (b) reveals the particles size distribution of undoped and KBM: $1.0Eu^{3+}$ phosphor are lie in a range of 0.2–0.6 μm [23].

3.5. Luminescence properties

Fig. 5(a) represents the photoluminescence excitation spectrum (PLE) for 1.0 mol% Eu^{3+} doped KBM phosphor in a range of 350–500 nm monitored by fixing the 612 nm emission wavelength [24]. Excitation spectrum exhibits various peaks located at 360, 380, 392, 413, 464 and 485 nm are attributed from $^7F_0$ ground state to distinct $^5D_4$, $^5L_7$, $^5L_6$, $^5D_3$, $^5D_2$ and $^5D_1$ excited states owing to f-f transitions of Eu^{3+} ions [25]. Two most prominent peaks were identified in near UV (n-UV) region at 392 nm and blue region at 464 nm wavelength resembles to $^7F_0 \rightarrow {}^5L_6$ and $^7F_0 \rightarrow {}^5D_2$ transitions, respectively [26]. The PLE measurement reveals that the n-UV/blue light can effectively be used to excite as-synthesized Eu^{3+} doped KBM phosphor.

Photoluminescence emission (PL) spectra for 1.0 mol% Eu^{3+} doped KBM phosphor have been measured at two discrete excitation wavelengths of 392 and 464 nm as shown in Fig. 5(b). PL spectra depict peaks located at 578, 588, 612, 657 and 704 nm in a range of 550–750 nm caused by the transition from lower excited state $^5D_0$ to various ground states as $^7F_0$, $^7F_1$, $^7F_2$, $^7F_3$ and $^7F_4$ levels, respectively [27]. Most significant peak in the PL spectra appeared at 612 nm ascribe to $^5D_0 \rightarrow {}^7F_2$ transition is electric dipole (ED) whereas peak positioned at 588 nm resembles to $^5D_0 \rightarrow {}^7F_1$ is magnetic dipole (MD) transition. In general, magnetic dipole transition follows selection rule $\Delta J = 1$ while electric dipole transition (hypersensitive transition) obeys selection rule $|\Delta L| \leq 2$; $|\Delta J| \leq 2$, and highly influenced by crystal field environment [28].

To identify the emission color of the as-synthesized phosphor, the Commission Internationale de l'Eclairage (CIE) coordinates are required to calculate through emission data. Fig. 6. depicts the CIE color coordinates of 1.0 mol% Eu^{3+} doped KBM phosphor have been found to be (0.657, 0.342) and (0.655, 0.344) under 392 and 464 nm excitation wavelengths, respectively. The CIE chromaticity coordinates are lie in the red region when exited with n-UV and blue light. The color correlated temperature (CCT) value of as-synthesized phosphor has been evaluated using McCamy's polynomial formula represented by equation [29]:

$$CCT = -437n^3 + 3601n^2 - 6861n + 5514.3 \quad (5)$$

where n represents the inverse slope line and equivalent to ratio of $(x\text{-}x_e)$ to $(y\text{-}y_e)$ and $x_e = 0.332$ and $y_e = 0.186$ are the epicentre of convergence. Moreover, the evaluated CCT value for 1.0 mol% Eu^{3+} doped KBM phosphor are 2901 and 2801 K under 392 and 464 nm excitation wavelength, respectively. The estimated CCT value of phosphor is less than 5000 K are much advantageous for warm light applications whereas value greater than 5000 K for cool light applications [30]. Therefore, the CCT values for 1.0 mol% Eu^{3+}

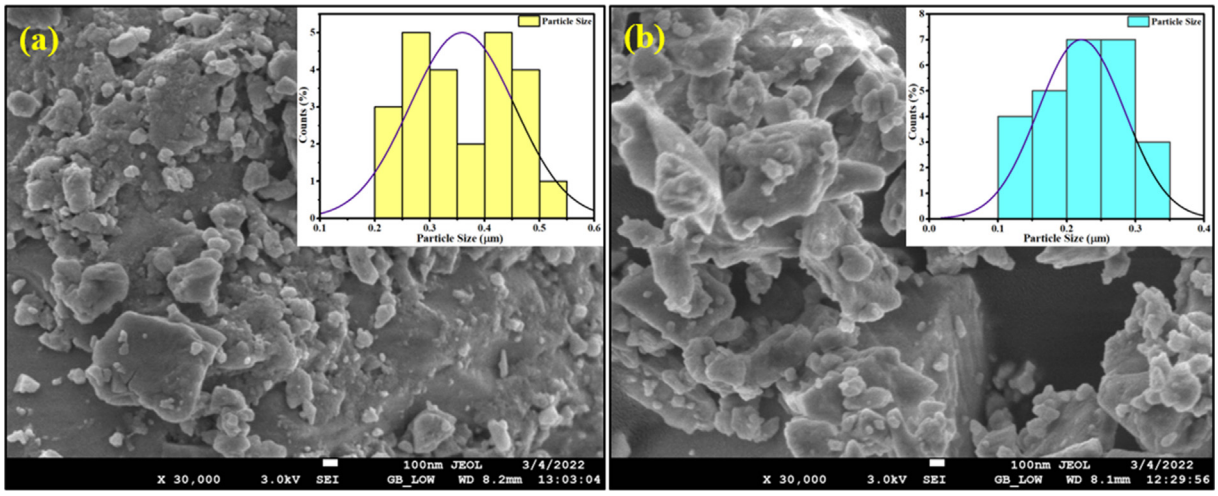


Fig. 4. (a) SEM micrographs of undoped KBM sample; (b) 1.0 mol% of Eu^{3+} doped KBM phosphor.

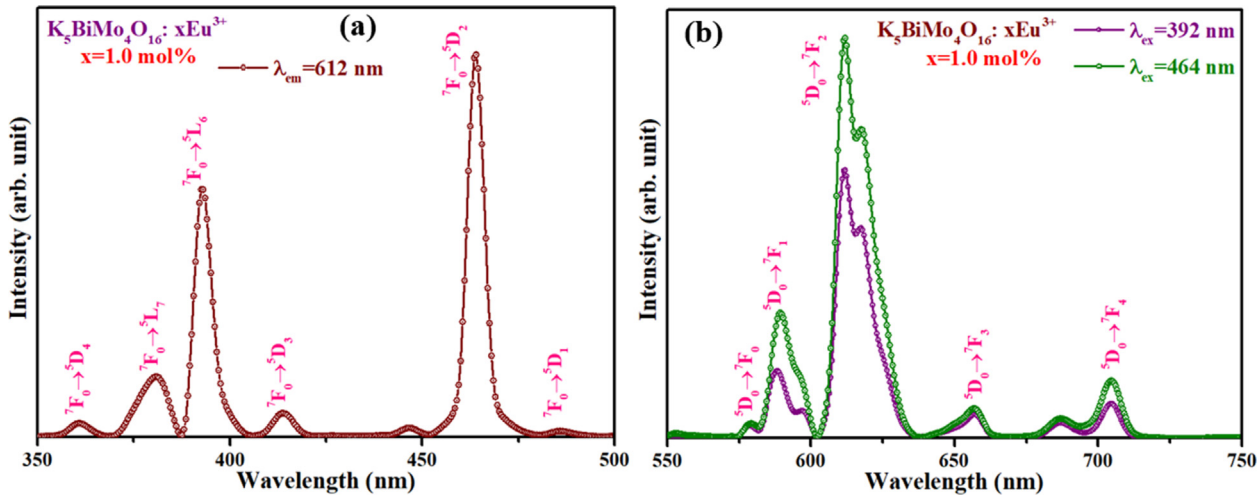


Fig. 5. (a) PLE spectrum of 1.0 mol% of Eu3^{+} doped KBM phosphor at 612 nm emission wavelength; (b) Comparison of Emission spectra of 1.0 mol% of Eu^{3+} doped KBM phosphor at two different excitation wavelengths i.e., 392 and 464 nm.

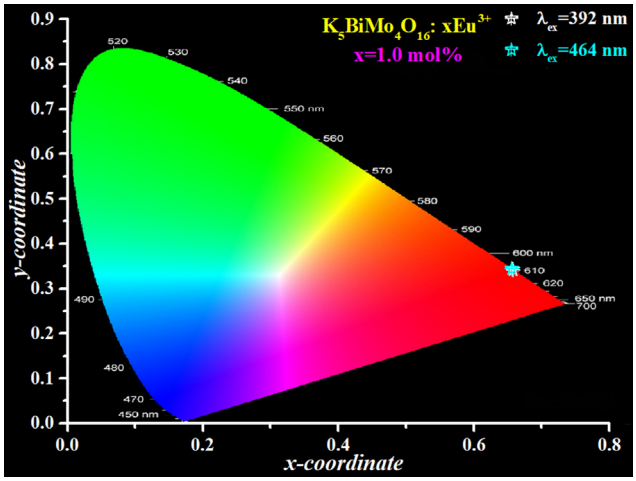


Fig. 6. CIE chromaticity diagram of 1.0 mol% of Eu^{3+} doped KBM phosphor under excitation wavelengths of λ_{ex} = 392 nm and λ_{ex} = 464 nm.

doped KBM phosphor are found to be less than 5000 K and can be promoted for warm w-LED applications.

4. Conclusions

Eu^{3+} doped red emitting $K_5BiMo_4O_{16}$ (KBM) phosphor has been prepared through solid state reaction route. The samples have been calcined at different temperatures (800, 900 and 1000 °C) and obtained the pure phase at 800 °C. Pure phase has been evident from the XRD pattern and compared with standard diffraction pattern (JCPDS card no. 29–0986). The crystallite sizes of as-synthesized undoped and Eu^{3+} doped KBM phosphors were calculated using Debye Scherrer formula and found to be 0.44 and 0.45 μm, respectively. The optical direct allowed band gap has been evaluated from K-M function of undoped and Eu^{3+} doped KBM phosphor are 3.60 and 3.58 eV, respectively. SEM micrograph displayed the agglomerated particles are lie in micro range. The CIE coordinates calculated based on emission spectra of 1.0 mol% Eu^{3+} doped KBM phosphor confirmed emission in red region, which are found to be (0.657, 0.342) and (0.655, 0.344) under 392 and

464 nm excitation wavelengths, respectively. The CCT value evaluated using McCamy's polynomial equation and found to be 2901 and 2801 K under 392 and 464 nm excitation wavelengths, respectively. Thus, the above-mentioned results highlighted that the phosphor can be effectively used for warm white light illumination and also shows great potential to contribute in a plethora of optoelectronic applications.

CRediT authorship contribution statement

Divyanshi Nagpal: Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **Tanisha Bhadauria:** Conceptualization, Methodology, Investigation, Data curation, Writing – original draft. **M. Jayasimhadri:** Supervision, Formal analysis, Visualization.

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

The authors are grateful to Ms. Deepali, Research Scholar, Luminescent Material Research Lab, Department of Applied Physics, Delhi Technological University for her help and support.

References

[1] Q.F. Wang, Y. Liu, Y. Wang, W. Wang, Y. Wan, G.G. Wang, Z.G. Lu, J. Alloys Compd. 625 (2015) 355–361.
[2] D. Lu, X. Gong, Y. Chen, J. Huang, Y. Lin, Z. Luo, Y. Huang, Opt. Mater. Express. 8 (2018) 259.
[3] G.R. Dillip, G.B. Kumar, V.R. Bandi, M. Hareesh, B. Deva Prasad Raju, S.W. Joo, L. K. Bharat, J.S. Yu, J. Alloys Compd. 699 (2017) 1108–1117.
[4] J. Lakde, C.M. Mehare, K.K. Pandey, N.S. Dhoble, S.J. Dhoble, J. Phys. Conf. Ser. 1913 (2021) 012029.
[5] S. Pradhan, H. Kaur, M. Jayasimhadri, Ceram. Int. 47 (2021) 27694–27701.
[6] B. Han, Y. Dai, J. Zhang, B. Liu, H. Shi, Ceram. Int. 44 (2018) 14803–14810.
[7] T. Sakthivel, G. Annadurai, R. Vijayakumar, X. Huang, J. Lumin. 205 (2019) 129–135.
[8] C. Guo, T. Chen, L. Luan, W. Zhang, D. Huang, J. Phys. Chem. Solids 69 (2008) 1905–1911.
[9] M. Sahu, N. Phatak, M.K. Saxena, RSC Adv. 11 (2021) 17488–17497.
[10] Z. Chen, H. Xu, C. Cao, X. Chen, M. Zhang, M. Jian, Y. Li, A. Xie, Coatings 11 (2021) 403.
[11] A. Xie, X. Yuan, Y. Shi, F. Wang, J. Wang, J. Am. Ceram. Soc. 92 (2009) 2254–2258.
[12] B. Han, J. Zhang, P. Li, J. Li, Y. Bian, H. Shi, Electron. Mater. 44 (2015) 1028–1033.
[13] L. Wang, X. Guo, Q. Song, J. Li, X. Wang, N. Wang, Y. Li, Y. Han, G. Jia, Mod. Phys. Lett. B. 32 (2018) 1850270.
[14] R.B. Deepali, H. Vandana, M.J. Kaur, J. Mater. Sci.: Mater. Electron. 32 (2021) 1650–1658.
[15] P. Jena, N. Nallamuthu, M. Venketswarulu, N. Satyanarayana, Ceram. Int. 40 (2014) 2349–2358.
[16] H. Kaur, M. Jayasimhadri, M.K. Sahu, P.K. Rao, N.S. Reddy, Ceram. Int. 46 (2020) 26434–26439.
[17] B. Amrithakrishnan, G. Subodh, Mater. Res. Bull. 93 (2017) 177–182.
[18] S. Dutta, S. Som, S.K. Sharma, RSC Adv. 5 (2015) 7380–7387.
[19] M. Tsvetkov, D. Elenkova, M. Milanova, Crystals 12 (2022) 120.
[20] J. Grigorjevaite, A. Katelnikovas, J. Mater. Res. Technol. 9 (2020) 15779–15787.
[21] Y. Zhai, W. Zhang, Y. Yin, J. Wang, C. Hu, X. Li, J. Mater. Sci.: Mater. Electron. 27 (2016) 9690–9698.
[22] C.S. Lim, A.S. Aleksandrovsky, M.S. Molokeev, A.S. Oreshonkov, D.A. Ikonnikov, V.V. Atuchin, Dalton Trans. 45 (2016) 15541–15551.
[23] M.K. Sahu, M. Jayasimhadri, J. Lumin. 227 (2020) 117570.
[24] H. Kaur, M. Jayasimhadri, Ceram. Int. 45 (2019) 15385–15393.
[25] Y. Shi, Z. Yang, W. Wang, G. Zhu, Y. Wang, Mater. Res. Bull. 46 (2011) 1148–1150.
[26] D. Shen, B. Zhao, S. Yang, S. Hu, J. Tang, X. Zhou, J. Yang, J. Mater. Sci. 52 (2017) 13868–13878.
[27] Y. Ding, Q. Meng, ChemistrySelect 4 (2019) 1092–1097.
[28] G. Rajkumar, V. Ponnusamy, G.V. Kanmani, M.T. Jose, Ceram. Int. 48 (2022) 10–21.
[29] D. Kumar Singh, K. Mondal, J. Manam, Ceram. Int. 43 (2017) 13602–13611.
[30] C. Dou, Z. Wang, Y. Yin, C. Liu, F. Zheng, S. Sun, B. Teng, J. Li, D. Zhong, J. Lumin. 239 (2021) 118326.



AKCE International Journal of Graphs and Combinatorics

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uakc20

Structural matrices for Signed Petri net

Payal & Sangita Kansal

To cite this article: Payal & Sangita Kansal (2022): Structural matrices for Signed Petri net, AKCE International Journal of Graphs and Combinatorics, DOI: 10.1080/09728600.2022.2070718

To link to this article: https://doi.org/10.1080/09728600.2022.2070718

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC

Published online: 10 May 2022.

Submit your article to this journal

Article views: 78

View related articles

View Crossmark data

Full Terms & Conditions of access and use can be found at
https://www.tandfonline.com/action/journalInformation?journalCode=uakc20

AKCE INTERNATIONAL JOURNAL OF GRAPHS AND COMBINATORICS
https://doi.org/10.1080/09728600.2022.2070718

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Structural matrices for Signed Petri net

Payal and Sangita Kansal

Department of Applied Mathematics, Delhi Technological University, Delhi, India

ABSTRACT

The developments in the field of Graph Theory and Petri net Theory in the form of balanceness and negative tokens respectively motivated the authors to bridge the gap between Petri net and Signed graph and introduce a new concept of Signed Petri net (SPN). In Petri net theory, matrices have been used to describe the structural behavior of the Petri net. Such matrices have been introduced for SPN which help in identifying relationships among the transitions and places of an SPN. Various subclasses of SPN are given along with characterizations of these subclasses using the matrices introduced. We consider ordinary SPNs (i.e. SPNs without multiple arcs) in the paper.

KEYWORDS

Signed graph; Signed Petri net; Petri net; state machine; free choice net; marked graph

1. Introduction

C.A. Petri first introduced Petri net (PN) in his dissertation submitted in 1962 [13]. Since then various extensions and results related to Petri net have been given by authors to simplify the structure of Petri net, to distinguish between its tokens, or to increase the modeling power of Petri net [1, 2, 4, 5, 8–11, 18–21]. A Signed Petri net (SPN) and techniques to analyze it have been introduced [7, 12]. SPNs are capable of modeling a large variety of systems and are preferred over a Petri net due to the presence of two types of tokens in it, positive and negative, which are distinguishable. Other advantages of using SPN over previously defined extensions of PN are the ability to assign a sign to vertices of an SPN which can be further utilized to introduce the concept of a balanced SPN.

Many authors in the field of Petri Net have utilized the concept of matrices to represent graphs and Petri Nets [16, 17] for computer implementation. Cantrell et al. used matrix reduction method for automatic verification of a model in Modeling and simulation [3]. Saha et al. utilized a relation between marking and firing vectors to obtain a simple and systematic technique of analysing Petri nets [14]. In [15], the author considers the pattern-type reachability analysis for the elastic dinning philosophers system based on Fractal Petri nets. Fractal algebra operations again implements by matrix operations.

Johnson et al. introduced several matrices which show the relationship between transitions and places in a PN [6]. This work inspired us to introduce such matrices for SPN which help in identifying the relation between its transitions and places. The matrix approach to SPN makes it easier to develop the algorithms related to various other concepts for SPN proposed by the authors. The structural matrices so introduced can be utilized for finding directed cycles in an SPN. The directed cycles will be utilized further by the authors for the introducing the concept of dynamical balanceness in their future work. The rest of the paper is structured as follows. In Sec. 2, three different matrices are defined by different products of the adjacency matrix of an SPN with its transpose and with itself. In fact, if all these matrices are given, the SPN structure can be obtained after analyzing them. In Sec. 3, subclasses of SPN are defined and it is shown how the matrices introduced in the previous section can be used to characterize these subclasses.

1.1. Signed Petri net [12]

Let $N = (P, T, I^-, I^+, \mu_0)$ be a PN. The arc set of the Petri net N is defined as:

$$E = \{(p,t) : I^-(p,t) > 0\} \cup \{(t,p) : I^+(p,t) > 0\}$$

A ***Signed Petri Net*** *(SPN)* is defined as a 3-tuple $N^* = (N', \sigma, \mu_0)$, where

(1) $N' = (P, T, I^-, I^+)$ is a Petri net structure.
(2) $\sigma : E \to \{+, -\}$, where E is the arc set of N'. An arc is called a positive or negative arc respectively according to the sign + or – assigned to it, using the function σ.
(3) $\mu_0 = (\mu_0^+, \mu_0^-)$ is the initial marking of SPN where
   (a) $\mu_0^+ : P \to \mathbb{N}$ gives the initial distribution of positive tokens in the places, called positive marking of SPN.
   (b) $\mu_0^- : P \to \mathbb{N}$ gives the initial distribution of negative tokens in the places, called negative marking of SPN.

CONTACT Sangita Kansal sangita_kansal15@rediffmail.com Department of Applied Mathematics, Delhi Technological University, Bawana Road, Delhi 110042 India.

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

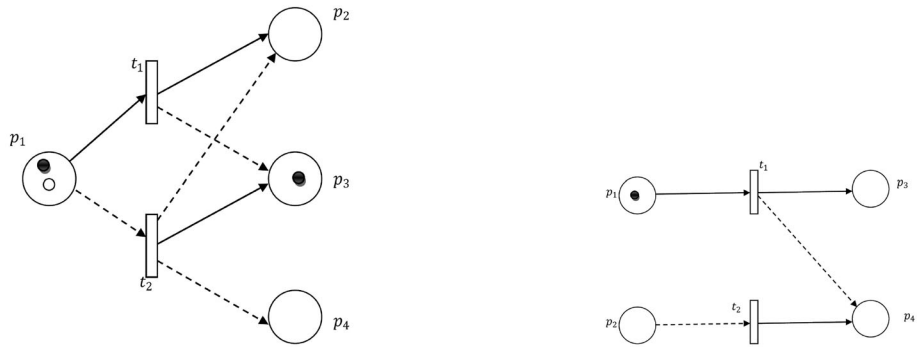


Figure 1. Execution of an SPN. (a) SPN with $\mu_0 = ((1, 0, 1, 0), (1, 0, 0, 0))$. (b) SPN with $\mu_0 = ((1, 0, 0, 0), (0, 0, 0, 0))$.

Thus, a marking in an SPN can be represented as a vector $\mu = (\mu^+, \mu^-)$ with $\mu^+, \mu^- \in \mathbb{N}^n, n = |P|$ such that ith component of μ is the value $\mu(p_i) = (\mu^+(p_i), \mu^-(p_i))$, which gives the number of positive and negative tokens at place p_i. Graphically, positive and negative arcs in an SPN are represented by solid and dotted lines respectively. A positive token is represented by a filled circle and a negative token by an open circle.

1.1.1. Execution rules for Signed Petri net

The execution of a Petri net depends on the distribution of tokens in the places of a Petri net. The execution takes place by firing of a transition. A transition may fire if it is enabled.

A transition t in an SPN N^* is *enabled* at a marking $\mu = (\mu^+, \mu^-)$ if

$$I^-(p,t) \leq \mu^+(p) \ \forall \ p \in {}^\bullet t \text{ for which } \sigma(p,t) = +$$
$$I^-(p,t) \leq \mu^-(p) \ \forall \ p \in {}^\bullet t \text{ for which } \sigma(p,t) = -$$

An enabled transition t may *fire* at $\mu = (\mu^+, \mu^-)$ provided $\exists\ p_k \in t^\bullet$ such that:

$$\sigma(t, p_k) = \begin{cases} + & \text{if } \sigma(p,t) = + \forall\ p \in {}^\bullet t \\ - & \text{if } \sigma(p,t) = - \forall\ p \in {}^\bullet t \\ +\text{ or }- & \text{if } \sigma(p,t) = + \text{for some } p \in {}^\bullet t \\ & \textit{and} - \text{for some } p \in {}^\bullet t \end{cases}$$

After firing, it yields a new marking $\mu_1 = (\mu_1^+, \mu_1^-)$ given by the rule:

$$\mu_1^+(p) = \mu^+(p) - I^-(p,t) + I^+(p,t) \ \forall p \in P$$
where (p, t) & (t, p) are positive arcs, if exist
$$\mu_1^-(p) = \mu^-(p) - I^-(p,t) + I^+(p,t) \ \forall p \in P$$
where (p, t) & (t, p) are negative arcs, if exist

We say that μ_1 is reachable from μ and write $\mu \xrightarrow{t} \mu_1$. We restrict the movement of positive (negative) tokens to positive (negative) arcs only.

A marking μ is reachable from μ_0 if there exists a firing sequence η that transforms μ_0 to μ and is written $\mu_0 \xrightarrow{\eta} \mu$. A *firing or occurence sequence* is a sequence of transitions $\eta = t_1 t_2 \ldots t_k$ such that

$$\mu_0 \xrightarrow{t_1} \mu_1 \xrightarrow{t_2} \mu_2 \xrightarrow{t_3} \mu_3 \ldots \xrightarrow{t_k} \mu$$

Note that a transition $t_j, 1 \leq j \leq k$ can occur more than once in the firing sequence η.

Let us look at the execution of an SPN with the help of an example.

In Figure 1(a), t_1 and t_2 both are enabled at μ_0. Firing of t_1 at μ_0 yields a new marking $\mu = ((0,1,1,0),(1,0,1,0))$ and firing of t_2 at μ_0 yields a new marking $\mu = ((1,0,2,0),(0,1,0,1))$. In Figure 1(b), t_1 is enabled, while t_2 is not. t_1 can fire at μ_0 to give a new marking $\mu = ((0,0,1,0),\ (0,0,0,1))$.

2. A matrix approach to SPN

2.1. Adjacency matrix

Since an SPN is a directed bipartite graph in which vertices (places and transitions) are connected (places with transitions and vice-versa) via two types of arcs: positive and negative, we can associate an adjacency matrix with an SPN. We define an *Adjacency Matrix* $A = [a_{ij}]$ of an SPN as a matrix of order $(m + n)$ where $|T| = m$ and $|P| = n$, with

$$a_{ij} = \begin{cases} 1^+, & \text{if there exists a directed positive arc} \\ & \quad \text{from vertex } i \text{ to vertex } j \\ 1^-, & \text{if there exists a directed negative arc} \\ & \quad \text{from vertex } i \text{ to vertex } j \\ 0\ , & \text{else} \end{cases} \quad (1)$$

Therefore, we have

$$A = \begin{bmatrix} 0 & B \\ C & 0 \end{bmatrix}$$

where $B = [b_{ij}]$ is a $m \times n$ matrix in which b_{ij} represents an arc (positive or negative), if exists, from transition t_i to place p_j and $C = [c_{ij}]$ is a $n \times m$ matrix in which c_{ij} represents an arc (positive or negative), if exists, to transition t_j from place p_i.

2.2. Conflict, confluence and precedence matrices

In order to introduce the conflict, confluence and precedence matrices of an SPN, we first define the necessary operations on the entries of the adjacency matrix.

- $1^+.1^+ = 1^+$
- $1^-.1^- = 1^-$
- $1^+.0 = 0 = 0.1^+$
- $1^-.0 = 0 = 0.1^-$

- $1^+.1^- = 1^* = 1^-.1^+$
- $\underbrace{1^i + 1^i \ldots + 1^i}_{k times} = k.1^i$, where $i \in \{+, -, *\}$, $k \in \mathbb{N}$
- $\underbrace{1^+ + 1^+ \ldots + 1^+}_{k_1 times} + \underbrace{1^- + 1^- \ldots + 1^-}_{k_2 times} + \underbrace{1^* + 1^* \ldots + 1^*}_{k_3 times} = k_1.1^+ + k_2.1^- + k_3.1^*$ where $k_1, k_2, k_3 \in \mathbb{N}$.

We form three different matrices by different products of the adjacency matrix of an SPN with its transpose and with itself as given below:

$$A'A = \begin{bmatrix} C'C & 0 \\ 0 & B'B \end{bmatrix} = \begin{bmatrix} \alpha & 0 \\ 0 & \beta \end{bmatrix} \quad (2)$$

$$AA' = \begin{bmatrix} BB' & 0 \\ 0 & CC' \end{bmatrix} = \begin{bmatrix} \delta & 0 \\ 0 & \gamma \end{bmatrix} \quad (3)$$

$$AA = \begin{bmatrix} BC & 0 \\ 0 & CB \end{bmatrix} = \begin{bmatrix} \phi & 0 \\ 0 & \psi \end{bmatrix} \quad (4)$$

Any entry of the matrices defined above gives two information: Magnitude and Symbol. The magnitude of an entry helps in identification of cardinality while, the symbols help to specify sign of arcs which is possible. For any entry $x = k.1^i$, $i \in \{+, -, *\}$, $k \in \mathbb{N} \cup \{0\}$, we call k as *magnitude* of x, denoted by $\|x\|$. $i \in \{+, -, *\}$ is called the *symbol* of a non-zero element x, which appears in the superscript of 1 and is denoted by $S(x)$. If $S(x) = +$, it represents a positive arc, if $S(x) = -$, it represents a negative arc. What $S(x) = *$ represents will be clear later in this section.

If $x = k_1.1^+ + k_2.1^- + k_3.1^*, k_1, k_2, k_3 \in \mathbb{N} \cup \{0\}$, then the magnitude of x is given by $\|x\| = k_1 + k_2 + k_3$ and if $k_i, k_j \neq 0$ for $i \neq j$ or $x = 0$ then $S(x)$ is not defined.

Definition 1. Transition-Conflict Matrix

The transition-conflict matrix is an $m \times m$ symmetric matrix given by, $C'C = \alpha = [\alpha_{ij}]$, where $\|\alpha_{ij}\| = |{}^\bullet t_i \cap {}^\bullet t_j| =$ the number of common input places of t_i and t_j and $\|\alpha_{ii}\| = |{}^\bullet t_i| =$ the total number of input places of t_i.

Definition 2. Place-Conflict Matrix

The place-conflict matrix is an $n \times n$ symmetric matrix given by, $B'B = \beta = [\beta_{ij}]$, where $\|\beta_{ij}\| = |{}^\bullet p_i \cap {}^\bullet p_j| =$ the number of common input transitions of p_i and p_j and $\|\beta_{ii}\| = |{}^\bullet p_i| =$ the total number of input transitions of p_i.

Definition 3. Transition-Confluence Matrix

The transition-confluence matrix is an $m \times m$ symmetric matrix given by, $BB' = \delta = [\delta_{ij}]$, where $\|\delta_{ij}\| = |t_i^\bullet \cap t_j^\bullet| =$ the number of common output places of t_i and t_j and $\|\delta_{ii}\| = |t_i^\bullet| =$ the total number of output places of t_i.

Definition 4. Place-Confluence Matrix

The place-confluence matrix is an $n \times n$ symmetric matrix given by, $CC' = \gamma = [\gamma_{ij}]$, where $\|\gamma_{ij}\| = |p_i^\bullet \cap p_j^\bullet| =$ the number of common output transitions of p_i and p_j and $\|\gamma_{ii}\| = |p_i^\bullet| =$ the total number of output transitions of p_i.

Definition 5. Transition-Precedence Matrix

The transition-precedence matrix is an $m \times m$ matrix given by, $BC = \phi = [\phi_{ij}]$, where $\|\phi_{ij}\| = |t_i^\bullet \cap {}^\bullet t_j| =$ the number of places that are in $t_i^\bullet$ as well as in ${}^\bullet t_j$ and $\|\phi_{ii}\| = |t_i^\bullet \cap {}^\bullet t_i| =$ the number of self-loops at t_i.

Definition 6. Place-Precedence Matrix

The place-precedence matrix is an $n \times n$ matrix given by, $CB = \psi = [\psi_{ij}]$, where $\|\psi_{ij}\| = |p_i^\bullet \cap {}^\bullet p_j| =$ the number of transitions that are in $p_i^\bullet$ as well as in ${}^\bullet p_j$ and $\|\psi_{ii}\| = |p_i^\bullet \cap {}^\bullet p_i| =$ the number of self-loops at p_i.

Let us consider an example and utilize the above mentioned matrices to find the relationships between transitions and places: The adjacency matrix for the SPN in Figure 2 is

$$A = \begin{bmatrix} 0 & 0 & 0 & 1^- & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1^- & 1^+ & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1^- \\ 1^+ & 0 & 0 & 0 & 0 & 0 & 0 \\ 1^- & 0 & 0 & 0 & 0 & 0 & 0 \\ 1^+ & 0 & 1^- & 0 & 0 & 0 & 0 \\ 0 & 1^- & 0 & 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & B \\ C & 0 \end{bmatrix}$$

We form the above mentioned six matrices for the SPN and make important observations.

(1) First we consider the product $A'A$, whose sub-matrices give us the transition-conflict and place-conflict matrices which are as:

$$\alpha = C'C = \begin{bmatrix} 2.1^+ + 1^- & 0 & 1^* \\ 0 & 1^- & 0 \\ 1^* & 0 & 1^- \end{bmatrix},$$

$$\beta = B'B = \begin{bmatrix} 1^- & 0 & 0 & 0 \\ 0 & 1^- & 1^* & 0 \\ 0 & 1^* & 1^+ & 0 \\ 0 & 0 & 0 & 1^- \end{bmatrix}$$

Now, looking at the transition-conflict matrix α and place-conflict matrix β, we can infer the following:

(a) Since $\alpha_{11} = 2.1^+ + 1^-$, we conclude that the transition t_1 has 3 input arcs, two positive and one of negative sign.

(b) $\alpha_{13} = 1^*$ tells us that there exists a common input place of t_1 and t_3. Since $S(\alpha_{13}) = S(1^*) = *$, this implies that one of the arc from the common input place to t_1 or t_3 is positive and the other negative.

(c) $\alpha_{22} = \alpha_{33} = 1^-$ shows that both t_2 and t_3 have one input arc. Also $S(\alpha_{22}) = S(\alpha_{33}) = S(1^-) = -$, which means that the input arc is negative in sign.

(d) $\beta_{11} = \beta_{22} = \beta_{44} = 1^-$ means that places p_1, p_2 and p_4 each have one input arc which is negatively signed.

(e) Since $\beta_{23} = 1^*$, there exists a common input transition of p_2 and p_3 and one of these input arcs from the common transition is positive and the other one negative.

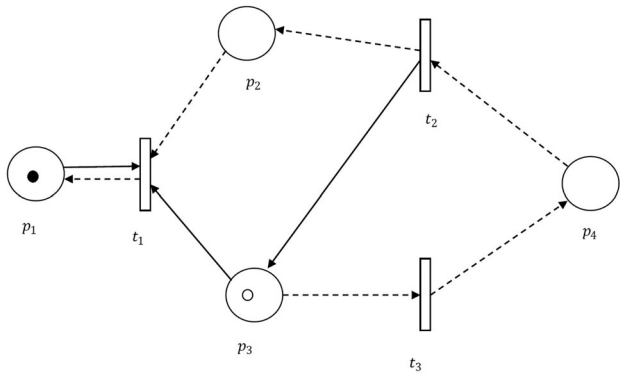


Figure 2. An SPN with initial marking ((1,0,0,0),(0,0,1,0)).

(2) Next, we look at the product AA', whose sub-matrices give us the transition-confluence and place-confluence matrices which are as:

$$\delta = BB' = \begin{bmatrix} 1^- & 0 & 0 \\ 0 & 1^+ + 1^- & 0 \\ 0 & 0 & 1^- \end{bmatrix},$$

$$\gamma = CC' = \begin{bmatrix} 1^+ & 1^* & 1^+ & 0 \\ 1^* & 1^- & 1^* & 0 \\ 1^+ & 1^* & 1^+ + 1^- & 0 \\ 0 & 0 & 0 & 1^- \end{bmatrix}$$

Similarly observing the transition-confluence matrix δ and place-confluence matrix γ, we can make the following observations:

(a) Since $\gamma_{33} = 1^+ + 1^-$, we conclude that the place p_3 has 2 output arcs, one positive and the other of negative sign.
(b) $\gamma_{21} = 1^*$ tells us that there exists a common output transition of p_1 and p_2 while the symbol $*$ implies that one of the arc to the common output transition is positive and the other negative.
(c) $\delta_{22} = 1^+ + 1^-$ shows that t_2 has two output arcs one of positive sign and the other of negative sign.
(d) $\delta_{11} = \delta_{33} = 1^-$ means that t_1, t_3 each have a negative output arc.

(3) Lastly, we check the product AA. The sub-matrices of AA give the transition-precedence and place-precedence matrices which are as:

$$\phi = BC = \begin{bmatrix} 1^* & 0 & 0 \\ 1^+ + 1^- & 0 & 1^* \\ 0 & 1^- & 0 \end{bmatrix},$$

$$\psi = CB = \begin{bmatrix} 1^* & 0 & 0 & 0 \\ 1^- & 0 & 0 & 0 \\ 1^* & 0 & 0 & 1^- \\ 0 & 1^- & 1^* & 0 \end{bmatrix}$$

Observing the transition-precedence matrix ϕ and place-precedence matrix ψ, we can make the observations given below:

(a) $\phi_{11} = 1^*$ means that transition t_1 has a self loop. The symbol $*$ implies that the self loop is of negative sign, i.e. one arc of self loop is positive while the other is negative.
(b) $\phi_{12} = 1^+ + 1^-$ tells us that there exists a common place which is output of t_1 and input of t_2. One of the arc is positive and the other negative.
(c) $\phi_{23} = 1^*$ shows that a common place is output of t_2 and input of t_3 with the arcs of opposite sign.
(d) $\psi_{12} = 1^-$ means that a transition exist which is output of p_1 and input of p_2, with each arc of negative sign.

Remark 1. Since, the proposed work is based on finding various structural matrices from the product of adjacency matrix with itself and with its transpose, therefore, the time complexity of the work is same as the time complexity of the matrix multiplication, which in our case is, $\mathcal{O}(N^3)$, where $N = m + n$.

The space complexity will be the complexity required to store a matrix, which is, $\mathcal{O}(n^2)$, where n is the order of the matrix.

3. Subclasses of SPNs

In Petri net theory, various subclasses of a Petri Net have been defined. In this section, such subclasses for an SPN have been introduced and it is shown how the matrices defined in the last section can be used to characterize these subclasses.

Definition 7. Signed State Machine

An SPN is called a Signed State Machine if each transition is restricted to have only one input and one output arc, i.e. $|^\bullet t| = |t^\bullet| = 1 \ \forall \ t \ \in T$ and both the input and output arcs are of same sign.

Theorem 1.

(i) *An SPN N^* is a signed state machine iff α_{ii}, $\delta_{ii} \in \{1^+, 1^-\}$ with $\alpha_{ii} = \delta_{ii} \ \forall \ i$.*
(ii) *If an SPN N^* is a signed state machine then both the place-conflict matrix, β and the place-confluence matrix, γ are diagonal matrices and the corresponding diagonal entries of transition-conflict matrix, α and transition-confluence matrix, δ are same i.e., $\alpha_{ii} = \delta_{ii} \ \forall \ i$.*
(iii) *If both β and γ are diagonal matrices and $\alpha_{ii} = \delta_{ii} \ \forall \ i$, then either N^* is a signed state machine or there exists a transition t such that $|^\bullet t| = 0$ or $|t^\bullet| = 0$.*

Proof.

(1) Let $\alpha_{ii}, \delta_{ii} \in \{1^+, 1^-\} \ \forall \ i \ \Rightarrow \|\alpha_{ii}\| = \|\delta_{ii}\| = 1 \ \forall i$. Therefore, $|^\bullet t_i| = |t_i^\bullet| = 1 \ \forall \ t$. Next, we have $\alpha_{ii} = \delta_{ii}$ which implies that the superscript of 1 is same, i.e. the input and output arcs of any transition t_i are of same sign. Thus, N^* is a signed state machine.
For the converse part, let N^* be a signed state machine. Since, $|^\bullet t_i| = |t_i^\bullet| = 1 \ \forall t_i \in T$, $\|\alpha_{ii}\| = \|\delta_{ii}\| = 1$. Also, by definition of a signed state machine, input and output arcs of a transition are of same sign, thus the

superscript of 1 in α_{ii}, δ_{ii} should be same for all i resulting in $\alpha_{ii} = \delta_{ii}$. Note that $\alpha_{ii}, \delta_{ii} \neq 1^*$, since then input and output arcs of transition t_i are of opposite sign. Therefore, $\alpha_{ii}, \delta_{ii} \in \{1^+, 1^-\}$

(2) N^* is a signed state machine, therefore for each transition t, $|{}^\bullet t| = |t^\bullet| = 1$. Thus $t \notin {}^\bullet p_i \cap {}^\bullet p_j$ *for* $i \neq j$ otherwise $p_i, p_j \in t^\bullet$ which is a contradiction to the definition of a signed state machine. Hence, $|{}^\bullet p_i \cap {}^\bullet p_j| = 0$ for $i \neq j$, which implies $\beta_{ij} = 0$ for $i \neq j$. Likewise, $t \notin p_i^\bullet \cap p_j^\bullet$ *for* $i \neq j$. Hence, $|p_i^\bullet \cap p_j^\bullet| = 0$ for $i \neq j$, implying $\gamma_{ij} = 0$ for $i \neq j$. Thus, both β and γ are diagonal matrices. Since the input and output arcs of a transition are of same sign, therefore, $\alpha_{ii} = \delta_{ii}\ \forall\ i$.

(3) If both β and γ are diagonal matrices, then $|{}^\bullet p_i \cap {}^\bullet p_j| = 0$ and $|p_i^\bullet \cap p_j^\bullet| = 0$ for $i \neq j$. Thus for any transition $t, |{}^\bullet t| = |t^\bullet| \leq 1$. Also, $\alpha_{ii} = \delta_{ii}\ \forall\ i$ implies that the input and output arcs of a transition are of same sign. Thus, either N^* is a signed state machine or $\exists$ a transition t such that either $|{}^\bullet t| = 0$ or $|t^\bullet| = 0$.

□

Definition 8. Signed Marked Graph

An SPN is called a signed Marked Graph if each place is restricted to have only one input and output arc, i.e. $|{}^\bullet p| = |p^\bullet| = 1\ \forall\ p\ \in P$ and both the input and output arcs are of same sign.

Theorem 2.

(1) *An SPN N^* is a signed marked graph iff $\beta_{ii}, \gamma_{ii} \in \{1^+, 1^-\}$ with $\beta_{ii} = \gamma_{ii}\ \forall\ i$.*

(2) *If an SPN N^* is a signed marked graph then both the transition-conflict matrix, α and the transition-confluence matrix, δ are diagonal matrices and the corresponding diagonal entries of place-conflict matrix, β and place-confluence matrix, γ are same i.e., $\beta_{ii} = \gamma_{ii}\ \forall\ i$.*

(3) *If both α and δ are diagonal matrices and $\beta_{ii} = \gamma_{ii}\ \forall\ i$, then either N^* is a signed marked graph or there exists a transition p such that $|{}^\bullet p| = 0$ or $|p^\bullet| = 0$.*

Proof.

(1) Let $\beta_{ii}, \gamma_{ii} \in \{1^+, 1^-\}\ \forall\ i \Rightarrow \|\beta_{ii}\| = \|\gamma_{ii}\| = 1\ \forall\ i$. Therefore, $|{}^\bullet p_i| = |p_i^\bullet| = 1\ \forall\ p_i \in P$. Next, we have $\beta_{ii} = \gamma_{ii}$ which implies that the superscript of 1 is same, i.e. the input and output arcs of any place p are of same sign. Thus, N^* is a signed marked graph.
The converse part follows from the definition of a signed marked graph.

(2) If N^* is a signed marked graph, then for each place p, $|{}^\bullet p| = |p^\bullet| = 1$. Thus $p \notin {}^\bullet t_i \cap {}^\bullet t_j$ *for* $i \neq j$ otherwise $t_i, t_j \in p^\bullet$ which is a contradiction to the definition of a signed marked graph. Hence, $|{}^\bullet t_i \cap {}^\bullet t_j| = 0$ for $i \neq j$. Likewise, $p \notin t_i^\bullet \cap t_j^\bullet$ *for* $i \neq j$. Hence, $|t_i^\bullet \cap t_j^\bullet| = 0$ for $i \neq j$. Thus, $\alpha_{ij} = \delta_{ij} = 0\ \forall i \neq j$ making both α and δ diagonal matrices. Next, by definition of signed marked graph $\beta_{ii} = \gamma_{ii}\ \forall\ i$.

(3) If both α and δ are diagonal matrices, then $|{}^\bullet t_i \cap {}^\bullet t_j| = 0$ and $|t_i^\bullet \cap t_j^\bullet| = 0$ for $i \neq j$. Thus, for any place p, $|{}^\bullet p| = |p^\bullet| \leq 1$. Also, $\beta_{ii} = \gamma_{ii}\ \forall\ i$ which implies input and output arcs of a place are of same sign. Thus, either N^* is a signed marked graph or $\exists\ p$ such that either $|{}^\bullet p| = 0$ or $|p^\bullet| = 0$.

Definition 9. Signed Free-Choice Net

An SPN is called a Signed free-choice net if

$$p_i^\bullet \cap p_j^\bullet \neq \emptyset \Rightarrow p_i^\bullet = p_j^\bullet \quad \& \quad |p_i^\bullet| = |p_j^\bullet| = 1$$

Remark 2. We define sets Z and Z' as:

$$Z = \{k_1.1^+ + k_2.1^- \ :\ k_1, k_2 \ \in\ \mathbb{N} \cup \{0\}\}$$
$$Z' = Z \backslash \{1^+, 1^-, 0\}$$

Theorem 3.

(1) *An SPN N^* is a signed free-choice net iff the following holds:-*
If $\alpha_{ii} \in Z'$, then $\alpha_{ij} = 0\ \forall\ i \neq j$. (If $\alpha_{ii} = 1^+$ or 1^-, then, $\alpha_{ij} \in \{0, 1^+, 1^-, 1^*\}$)

(2) *An SPN is a signed free-choice net iff $\gamma_{ij} \neq 0$ for some i and $j \Rightarrow \|\gamma_{ij}\| = \|\gamma_{ii}\| = \|\gamma_{jj}\| = 1$.*

Proof.

(1) Suppose N^* is a signed free-choice net. If $p_i^\bullet \cap p_j^\bullet \neq \emptyset$, then $p_i^\bullet = p_j^\bullet = \{t\}$ for some $t \in T$. Now $\alpha_{ii} \in Z'$ implies $\alpha_{ii} = k_1.1^+ + k_2.1^-$ for some non-zero $k_1, k_2 \in \mathbb{N} \Rightarrow \|\alpha_{ii}\| = k_1 + k_2 = |{}^\bullet t_i|$. Let $p_1, p_2, \ldots p_{k_1+k_2} \in {}^\bullet t_i \Rightarrow t_i \in p_1^\bullet \cap p_2^\bullet \cap \ldots \cap p_{k_1+k_2}^\bullet$ i.e., $\cap_{i=1}^{k_1+k_2} p_i^\bullet \neq \emptyset \Rightarrow p_1^\bullet = p_2^\bullet = \ldots = p_{k_1+k_2}^\bullet = \{t_i\}$, i.e., for every $p \in {}^\bullet t_i$, $p^\bullet = \{t_i\}$. Therefore, $\|\alpha_{ij}\| = |{}^\bullet t_i \cap {}^\bullet t_j| = 0$, since any p in ${}^\bullet t_i$ cannot be in ${}^\bullet t_j$, otherwise $\{t_i, t_j\}$ would be contained in $p^\bullet = \{t_i\}$. Therefore, if $\alpha_{ii} \in Z'$, then $\alpha_{ij} = 0\ \forall\ i \neq j$. Note that $\|\alpha_{ij}\| \leq \|\alpha_{ii}\|\ \forall\ i, j$ (since ${}^\bullet t_i \cap {}^\bullet t_j \subseteq {}^\bullet t_i$). Thus, if $\alpha_{ii} \in \{1^+, 1^-\}$ then, $\alpha_{ij} \in \{0, 1^+, 1^-, 1^*\}$.
Conversely, suppose that $\alpha_{ii} \in Z'$, then $\alpha_{ij} = 0\ \forall\ i \neq j$ i.e., if $|{}^\bullet t_i| > 1$, then $|{}^\bullet t_i \cap {}^\bullet t_j| = 0$. We need to show that N^* is a signed free-choice net. For this, we assume that $p_i^\bullet \cap p_j^\bullet \neq \emptyset$ for $i \neq j$. Let $t_k, t_l \in p_i^\bullet \cap p_j^\bullet$. Then, ${}^\bullet t_k$ contains $\{p_i, p_j\}$ but so does ${}^\bullet t_l$, a contradiction (since $\alpha_{ij} = 0$). Thus, $p_i^\bullet \cap p_j^\bullet$ can contain atmost one transition and $p_i^\bullet = p_j^\bullet = t$ for some $t \in T$. Hence, N^* is a signed free-choice net.

(2) Suppose N^* is a signed free-choice net. If $\gamma_{ij} \neq 0$, then $|p_i^\bullet \cap p_j^\bullet| \neq 0\ \forall i \neq j$ i.e., $p_i^\bullet \cap p_j^\bullet \neq \emptyset \Rightarrow p_i^\bullet = p_j^\bullet$ and $|p_i^\bullet| = 1 = |p_j^\bullet|$ (Since, N^* is a signed free-choice net). Therefore, $\|\gamma_{ii}\| = \|\gamma_{jj}\| = 1$. Now, $\|\gamma_{ij}\| = |p_i^\bullet \cap p_j^\bullet| \neq 0$. Also, $|p_i^\bullet \cap p_j^\bullet| \leq |p_i^\bullet|$ (As $p_i^\bullet \cap p_j^\bullet \subseteq p_i^\bullet$). So, $|p_i^\bullet \cap p_j^\bullet| = 1 = \|\gamma_{ij}\|$. Hence, $\|\gamma_{ij}\| = \|\gamma_{ii}\| = \|\gamma_{jj}\| = 1$.
Conversely, we need to show that N^* is a signed free-choice net. Let $p_i^\bullet \cap p_j^\bullet \neq \emptyset$, therefore, $|p_i^\bullet \cap p_j^\bullet| \neq 0$. By hypothesis, $\|\gamma_{ij}\| = \|\gamma_{ii}\| = \|\gamma_{jj}\| = 1$. Therefore, $|p_i^\bullet| = |p_j^\bullet| = |p_i^\bullet \cap p_j^\bullet| = 1$. Let $p_i^\bullet \cap p_j^\bullet = \{t\}$ for some $t \in T$. Therefore, $p_i^\bullet \cap p_j^\bullet = p_i^\bullet = p_j^\bullet = \{t\}$(As $p_i^\bullet \cap$

$p_j^\bullet \subseteq p_i^\bullet$ & $p_i^\bullet \cap p_j^\bullet \subseteq p_j^\bullet$). Hence, N^* is a signed free-choice net.

Definition 10. Extended Signed Free-Choice Net

An SPN is called an Extended Signed Free-choice net if

$$p_i^\bullet \cap p_j^\bullet \neq \emptyset \Rightarrow p_i^\bullet = p_j^\bullet$$

Theorem 4. *An SPN N^* is an extended signed free-choice net iff $\gamma_{ij} \neq 0$ for some i and $j \Rightarrow \|\gamma_{ii}\| = \|\gamma_{jj}\| = \|\gamma_{ji}\|$.*

Proof. Suppose N^* is an extended signed free-choice net. Let $\gamma_{ij} \neq 0$ therefore, $|p_i^\bullet \cap p_j^\bullet| \neq 0 \Rightarrow p_i^\bullet \cap p_j^\bullet \neq \emptyset$ which implies $p_i^\bullet = p_j^\bullet$ so, $p_i^\bullet = p_i^\bullet \cap p_j^\bullet$. Thus, $\|\gamma_{ii}\| = \|\gamma_{jj}\| = \|\gamma_{ji}\|$.

Conversely, suppose that $\gamma_{ij} \neq 0 \Rightarrow \|\gamma_{ii}\| = \|\gamma_{jj}\| = \|\gamma_{ji}\|$, i.e., $p_i^\bullet \cap p_j^\bullet \neq \emptyset \Rightarrow \|\gamma_{ij}\| = \|\gamma_{ii}\|$. Therefore, $|p_i^\bullet \cap p_j^\bullet| = |p_i^\bullet|$ then, $p_i^\bullet = p_i^\bullet \cap p_j^\bullet$. Likewise, $\|\gamma_{ij}\| = \|\gamma_{jj}\|$ so that $p_j^\bullet = p_i^\bullet \cap p_j^\bullet$ implying $p_i^\bullet = p_j^\bullet$. Thus, N^* is an extended signed free-choice net. □

Definition 11. Signed Simple Net

An SPN is called a Signed Simple net if

$$p_i^\bullet \cap p_j^\bullet \neq \emptyset \Rightarrow \text{either} \quad p_i^\bullet \subset p_j^\bullet \quad \text{or} \quad p_j^\bullet \subset p_i^\bullet$$

Theorem 5. *An SPN N^* is a signed simple net iff $\gamma_{ij} \neq 0$ for some i and $j \Rightarrow$ either $\|\gamma_{ij}\| = \|\gamma_{ii}\|$ or $\|\gamma_{ij}\| = \|\gamma_{jj}\|$.*

Proof. Suppose that N^* is a signed simple net. If $\gamma_{ij} \neq 0 \Rightarrow p_i^\bullet \cap p_j^\bullet \neq \emptyset$ then, either $p_i^\bullet \subset p_j^\bullet$ or $p_j^\bullet \subset p_i^\bullet \Rightarrow$ either $p_i^\bullet \cap p_j^\bullet = p_i^\bullet$ or $p_i^\bullet \cap p_j^\bullet = p_j^\bullet$ which implies either $\|\gamma_{ij}\| = \|\gamma_{ii}\|$ or $\|\gamma_{ij}\| = \|\gamma_{jj}\|$.

The converse can be obained by reversing the steps. □

Remark 3. The space complexity of finding any of the subclasses of SPN is same as the space compexity for storing a matrix, i.e., $\mathcal{O}(n^2)$, where n is the order of the matrix.

The time complexity of finding signed state machine and signed marked graph is $\mathcal{O}(n)$ where n is the order of the matrix, since in these cases we only consider the diagonal entries of the matrices concerned. The time complexity of finding rest of the subclasses is $\mathcal{O}(n^2)$, where n is the order of the matrix because in these cases all the elements of the concerned matrix are considered.

4. Conclusion

As the graphs/images are unsuitable for computer implementation and poses problems in storage since images use more memory as compared to text, a matrix approach for SPN has been introduced. Hence, using the matrices defined in this paper we can store the structural information of the SPN and utilize these matrices to find relationships between transitions and places of an SPN. The characterizations for subclasses of SPN are given using the matrices defined in the paper.

For future scope, the matrices introduced in the paper can be used to find directed cycles in an SPN, which can then be used for defining the concept of dynamical balanceness for an SPN.

References

[1] Alla, H., David, R. (1998). Continuous and hybrid Petri Nets. *J. Circuit Syst. Comp*. 08(01): 159–188.

[2] Bartoletti, M., Cimoli, T., Pinna, G. M. (2015). Lending Petri nets. *Sci. Comput. Program*. 112: 75–101.

[3] Cantrell, W. A., Maxwell, K. P., Petty, M. D., Whitaker, T. S. (2020). Matrix reduction verification of extended Petri nets. *Proceedings of the 2020 ACM Southeast Conference, Association for Computing Machinery, New York, NY*, pp. 296–299.

[4] Chen, S.-M., Ke, J.-S., Chang, J.-F. (1990). Knowledge representation using fuzzy Petri nets. *IEEE Trans. Knowl. Data Eng*. 2(3): 299–309.

[5] Du, Y. Y., Guo, B. Q. (2009). Logic Petri nets and equivalency. *Inf. Technol. J*. 8(1): 95–100.

[6] Johnson, J. L., Murata, T. (1985). Structure matrices for Petri nets and their applications. *J. Franklin Inst*. 319(3): 299–309.

[7] Kansal, S., Dabas, P. (2021). An introduction to Signed Petri net. *J. Math*. 2021: 1–8.

[8] Kansal, S., Singh, G. P., Acharya, M. (2010). On the Petri nets generating all the binary n-vectors. *Sci. Math. Jpn*. 71(2): 209–216.

[9] Kansal, S., Singh, G. P., Acharya, M. (2011). 1-safe Petri nets generating every binary n-vector exactly once. *Sci. Math. Jpn*. e-2011: 127–134.

[10] Marsan, M. A. (1989). Stochastic Petri nets: An elementary introduction. In *Advances in Petri Nets. LNCS*, Vol. 424. Springer, pp. 1–30.

[11] Murata, T. (1989). Petri nets: Properties, analysis and applications. *Proc. IEEE*. 77(4): 541–580.

[12] Payal, P., Kansal, S. (2020). Analysis of a Signed Petri net. *Int. J. Comput. Sci. Math*. (Accepted).

[13] Petri, C. A. (1962). *Kommunikation Mit Automaten*. Bonn: Schriften des Institutes fur Instrumentelle Mathematik.

[14] Saha, B., Bandyopadhyay, S. (1988). Representation and analysis of petrinets via the matrix state equation approach. *Int. J. Electron*. 65(1): 1–7.

[15] Semenov, A. S. (2020). Matrix approach to pattern-type reachability analysis of fractal Petri nets. In *2020 7th International Conference on Control, Decision and Information Technologies (CoDIT)*, Vol. 1. pp. 211–216.

[16] Spiteri Staines, T. (2016). Modelling simple network graphs using the matrix vector transition net. *Int. J. Comput. Commun*. 10: 11–17.

[17] Staines, S. A. (2019). Alternative matrix representation of ordinary petri nets. *WSEAS Trans. Comput*. 18: 11–18.

[18] Valk, R. (2004). *Object Petri Nets*. LNCS 3098, Lecture on Concurrency and Petri Nets. Berlin, Heidelberg: Springer, pp. 819–848.

[19] van der Aalst, W. M. P. (1993). *Multi-Dimensional Petri* Nets. Computing Science Notes Eindhoven, Vol. 9326. The Netherlands: Technische Universiteit Eindhoven.

[20] van der Aalst, W. M. P. (1998). The application of Petri nets to workflow management. *J. Circuit Syst. Comp*. 08(01): 21–66.

[21] Wang, J. (1998). Time Petri nets. In: *Timed Petri Nets*. The Kluwer International Series on Discrete Event Dynamic Systems, Vol. 9. Boston MA: Springer, pp. 63–123.

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

STUDY ABOUT MACHINE OPERATIONS AND ITS DIFFERENT OPERATIONS

Chander Mohan Aggarwal[*1], Bhavya Dhanwaria[*2], Dr. Ravi Butola[*3], Prof. A.K. Madan[*4]

[*1,2,3,4]Department Of Mechanical Engineering, Delhi Technological University, Delhi, India.

ABSTRACT

A machine tool is a machine tool used to shape and grind metals or other materials, usually by cutting, drilling, grinding or shearing. There are cases where individual parts of a composite unit are completed in a series of processes that may include various cuts, grinding, plastic modification, heat treatment etc. In this we have explained about the machining and different machine tool like Grinding, Drilling, Shearing etc.

Keywords: Turning, Lathe, Milling, Shaper, Drilling.

I. INTRODUCTION

The development of machine tools for manufacturing is closely associated with the Industrial Revolution, marked by James Watt's successful development of the steam engine in 1776. When he began to design steam engines in 1763, his main challenge was to build cylinders. Drill precisely enough to prevent steam from escaping around the piston. The original cylinder was made of sheet metal, certainly not working properly. The problem was finally solved by John Wilkinson when he invented the horizontal water wheel reamer around 1775. This machine is often considered the first machine tool.

Machining is a general term describing a group of processes that include removing material and modifying the surface of a part after it has been manufactured by various methods. So machining includes secondary and finishing operations. The basic principle of all machining is to provide adequate relative movements between the tool and the workpiece to produce a characteristic geometry. The three main machining processes are classified into turning, drilling and milling.

Turning

Turning is a machining process in which a cutting tool, usually a non-rotating tool, describes a twisted toolpath by moving more or less linearly as the part rotates. The turning process generally involves rotating a part while a single point tool is moved parallel to the axis of rotation. The turning processes are typically carried out on a lathe, considered to be the oldest of machine tools.

Turning specific operations include:

1. Tapered turning
2. Spherical generation
3. Hard turning
4. Facing

Drilling and Milling

Drilling is the act of making a round hole by removing a mass of metal from a product with a cutting tool called a drill. A drill is a rotary head cutting tool with one or more cutting lips and usually one or more flutes for chip transfer and cutting fluid reception. A punch drill is a machine tool designed to drill holes in metal. It is one of the most versatile and important machine tools in the workshop.

This machine consists of following parts

1. Base
2. Pillar
3. Main drive
4. Drill spindle
5. Feed handle
6. Work table

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:04/April-2022 Impact Factor- 6.752 www.irjmets.com

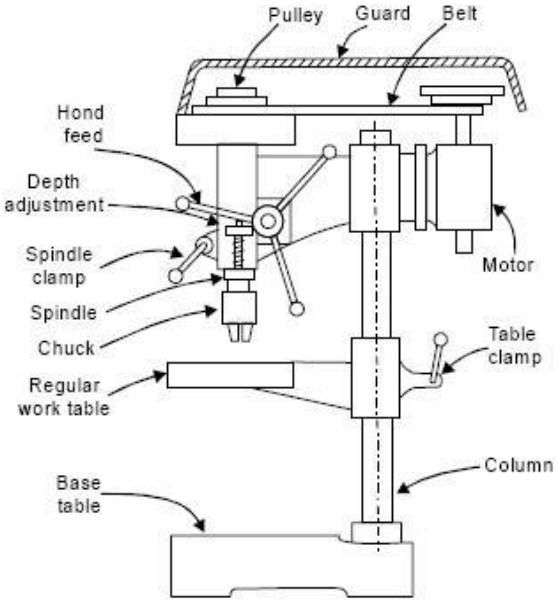


Figure 1: Diagram of Drilling Machine.

Milling is a machining process that uses a rotary cutter to remove material by passing a cutter through the workpiece. Milling covers a wide variety of operations and machines ranging from small individual parts to large group milling operations. It is one of the most widely used processes for machining custom parts to exact tolerances.

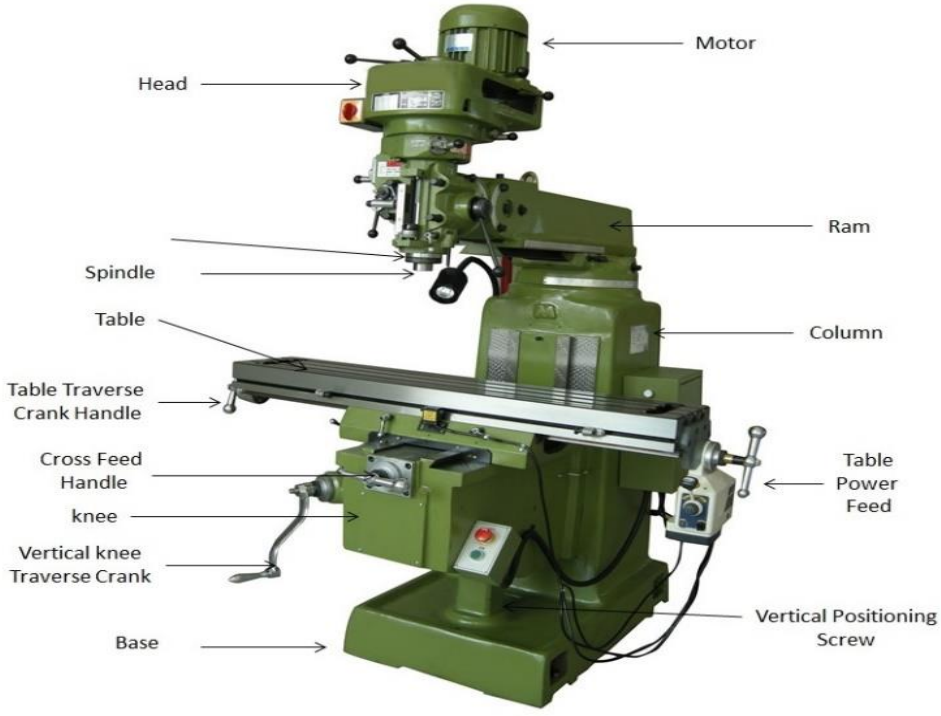


Figure 2: Diagram of Milling Machine.

II. RESULTS AND DISCUSSION

We have discussed the different types of machine operation by specially focusing in Turning, Grinding and milling. We found the different applications of these machine operations in different industries. As milling machine can hold single as well as multiple cutter while working at a same time, this become the reason of its importance in production work. Similarly, Turing machines find applications in complexity research and algorithmic information theory, software testing, high-performance computing, machine learning, software engineering, computer networks, and evolutionary computing.

III. CONCLUSION

Machining is the production of a material into a desired shape and size. Usually, machining refers to metal processing; however, it can also be used to refer to the production of wood, plastic, ceramic, stone and other materials. Machining is a necessary process for many industries, especially those related to construction and manufacturing. Because machining is so versatile and precise, it allows businesses and industries to design products and tools that meet their exacting needs.

ACKNOWLEDGEMENTS

We wish to express our gratitude to all those who provided help and cooperation in various ways at the different stages for this research project. Also, we would like to express our sincere appreciation to our teachers Prof. A.K. Madan and Prof. Dr Ravi Butola. Also, we would like to thank our friends and special thanks to Abhishek Kishore for motivating us.

IV. REFERENCES

[1] R. Woodbury, "Studies in the History of Machine Tools", MIT Press, 1972.

[2] S. Bahaley, A. Awate, A. Saharkar, "Performance Analysis of Pedal Powered Multipurpose Machine", International Journal of EngineeringResearch and Development (IJERD), Vol.1, Issue.5, ISSN: 2278-0181, 2012.

[3] K. Patel, N. Kumar, S. Agrawal, "Design and development of a manual Multifunctional portable machine" International Journal ofinnovations in Engineering and technology (IJIET) Vol.2 Issue 2, ISSN: 2319-1058 page 157-163, 2013.

[4] S. Srivastava, S. Shivam, C. Khatri, "Multi-Function Operating Machine: A Conceptual Model", IOSR Journal of Mechanical and CivilEngineering (IOSR-JMCE) e-ISSN: 2278-1684, p-ISSN: 2320-334X, Volume 11, Issue 3 Ver. III , PP 69-75, 2014.

[5] Linxu, W. Bai, R. Jingyu, Q. Li., "Design and Implementation of the Reciprocating Pedal Powered Electricity Generating Device", AdvancedMaterials Research ,Vol.282-283 ,pp 735-738, 2011.

[6] T. Moriwaki, "Trends in Recent Machine Tool Technologies", NTN Technical Review No.74, 2006.

[7] T. Moriwaki, "Multi-functional machine tool", CIRP Annals - Manufacturing Technology DOI: 10.1016/j.cirp., 2008.

[8] Y. Koren, U. Heisel, F. Jovane , T. Moriwaki, G. Pritschow, G. Ulsoy, H. VanBrussel, "Reconfigurable Manufacturing Systems", Annals ofthe CIRP 48(2):527–540, 1999.

[9] R.G.Landers, HB. Min, Y. Koren, "Reconfigurable Machine Tools", Annals of the CIRP 50(1):269–274, 2001.

IJE TRANSACTIONS A: Basics Vol. 35, No. 07, (July 2022) 1355-1364

International Journal of Engineering

Journal Homepage: www.ije.ir

Surface Roughness in Abrasive Mixed Rotary Electrical Discharge Machining: Experimental Approach

A. Lamba*[a], V. Vipin[b]

[a] *Department of Mechanical and Automation Engineering, G. B. Pant Okhla-I Campus, DSEU, Delhi, India*
[b] *Department of Mechanical Engineering, Delhi Technological University, Delhi, India*

PAPER INFO

Paper history:
Received 22 January 2022
Received in revised form 19 March 2022
Accepted 30 March 2022

Keywords:
Electrical Discharge Machining
D3 Steel
Full Factorial Experimentation
Surface Roughness
Analysis of Variance

ABSTRACT

This research study focused on the alliance of blending silicon carbide powder of size 60 µm into dielectric oil (kerosene) with electrode revolution in electrical discharge machining of D3 steel using electrode made of copper. Peak current, pulse on-time, and electrode revolution were picked as machining variables to check their influence on the surface roughness of the workpiece. The full factorial experimentation was performed with 27 experimental runs. The analysis of variance result indicated electrode revolution as the most influential variable with a percentage aid of 32.75%, followed by peak current and pulse on-time, which accounted for 28.98 and 6.09%, respectively. The decrement in surface roughness was recorded at higher electrode rotational speed. The field emission scanning electron microscopy of machined samples was performed and results showed significant improvement on surface characteristics. The surface finish improved from 12.36 µm to 10.32 µm in silicon carbide powder blended rotational electrical discharge machining over traditional electrical discharge machining. The minimum surface roughness achieved was 4.81 µm at optimal conditions.

***doi**: 10.5829/ije.2022.35.07a.14*

NOMENCLATURE

EDM	Electrical discharge machining	SR	Surface roughness
Pc	Peak current (A)	SiC	Silicon carbide
Pot	Pulse on-time (µs)	ANOVA	Analysis of variance
Erpm	Electrode revolution (rpm)	FESEM	Field emission scanning electron microscopy

1. INTRODUCTION

The ever increasing demand for highly accurate cutting methods has led to the origination of electrical discharge machining (EDM). The material removal in EDM happens with electrical sparks enabling the ejection of tiny debris particles from the workpiece surrounded by EDM fluid. As the sparks generate beneath the electrode's surface, the impression generated on the workpiece's surface replicates the shape of the electrode. The temperature generated by the sparks is of the order of 8000-20000 ℃ which evaporates the material [1]. The debris thus generated are removed with the help of circulating oil. This cutting style has completely revolutionized the traditional approach with its extraordinary features like contact-less machining, the ability to cut intricate shapes of material of high hardness including composites with a softer tool material. This method has found a concrete place in almost all machining industries throughout the world. Apart from its many advantages, the process also suffers from disadvantages. The disadvantages include limited material removal capability, difficulty in producing sharp corners in complex shapes, rough surfaces, etc. [2]. In the quest for improvements over these limitations, the researchers initially focused on optimal combinations of input process parameters, and thereafter, as time passed, the focus was shifted to the development of different

*Corresponding author emial: ajay.lamba@dseu.ac.in (A. Lamba)

Please cite this article as: A. Lamba, V. Vipin, Surface Roughness in Abrasive Mixed Rotary Electrical Discharge Machining: Experimental Approach, ***International Journal of Engineering, Transactions A: Basics***, Vol. 35, No. 07, (2022) 1355-1364

EDM variants. In the past two decades, researchers [3-10] all over the world tailored various successful attempts by their profound search in the domain of electrical discharge machining. These successful attempts by researchers led to the development of different variants resulting in near dry EDM, powder blended EDM, EDM with electrode revolution, vibration-assisted EDM, ultrasonic vibration-assisted EDM, ultrasonic vibration-assisted EDM with a cryogenically cooled electrode, magnetic field-assisted EDM, rotating magnetic field and ultrasonic-assisted EDM, etc. Out of the various research fields, two areas namely, EDM with electrode revolution and powder blended EDM were focused for deeper analysis.

2. LITERATURE SURVEY

The literature review is summarized in Table 1. The presented literature [11-13] showed that surface roughness is improved by electrode revolution [4, 14-17]. It can be seen that the blending of suitable powder type into main dielectric media improved the surface roughness during machining with EDM. In light of these observations [4, 11-17], it was seen that the individual studies are available that improved the same output characteristic i.e. surface roughness. Hence, a research gap was identified by clubbing these two approaches intending to improve the surface roughness even more. Hence, a study involving both electrode revolution and powder blending in the same setup was planned to assess the surface roughness for a work-tool combo of D3 steel

TABLE 1. Summary of literature review

Author	EDM type	W/p material	Tool material	Process parameters	Output responses	Methodology	Findings	Ref.
Chattopadhyay et al.	Rotary EDM	EN8 steel	Cu	peak current, pulse on-time, tool rotation	SR, MRR, EWR	TRD	The study revealed peak current and electrode revolution as highly influential factors affecting SR.	[5]
Shih and Shu	Rotary EDM	AISI D2 tool steel	Cu	discharge depth, electrode rotational speed, discharge direction, polarity, pulse duration, flushing direction, dielectric fluid	SR, MRR, EWR	TRD	1. They claimed minimum surface roughness of 2 µm (R_t) with a negative polarity electrode. 2. Pulse duration and peak current were reported to be the most influential parameters.	[11]
Aliakbari and Baseri	Rotary EDM	X210Cr12 tool steel	Cu	pulse on time, peak current, electrode rotation	SR, MRR, EWR, overcut	TRD	1. They achieved a decreasing trend in the value of SR with the increase in electrode rpm. 2. The minimum surface roughness recorded with a solid electrode, one-eccentric hole electrode, and two-eccentric holes electrode were 5.755 µm, 6.209 µm and 6.68 µm, respectively.	[12]
Dwivedi and Choudhury	Rotary EDM	AISI D3 tool steel	Cu	tool rotation, current	SR	–	1. Electrode revolution caused fewer overall micro-cracks and resulted in a better surface. 2. The improvements in surface characteristics were 10% higher in comparison to stationary EDM i.e. without electrode revolution.	[13]
Kansal et al.	Silicon powder blended EDM	EN31 tool steel	Cu	powder concentration, duty cycle, pulse on-time, peak current	SR, MRR	RSM	1. The apparent improvement in surface characteristics was seen with the use of silicon powder in dielectric oil. 2. SR was mainly affected by peak current and concentration of powder. 3. The lowest SR value was reported at high peak current and powder concentration.	[4]

Bhattacharya et al.	Gr and Al powder blended EDM	EN 31, H 11, HCHCr die steel;	Cu, W-Cu	pulse on-time, pulse off-time, current, powder, electrode, dielectric, w/p material	SR, MRR, TWR	TRD	1. Powder along with other factors such as pulse on-time and peak current significantly affected SR. 2. The surface quality with graphite powder blended oil was found better.	[14]
Bhattacharya et al.	Si, Gr, and W powder blended EDM	HCHCr steel, H 11 hot die steel, AISI 1045 steel;	Gr, Br, W-Cu	pulse on-time, current, type of powder, powder concentration	SR, microhardness	TRD	1. The results indicated the improvement in the surface finish of the processed parts. 2. The surface finish was mostly affected by pulse on-time, peak current, and powder concentration.	[15]
Singh et al.	Gr powder blended EDM	Super Co 605;	Gr	pulse on-time, pulse off-time, current, polarity, discharge voltage, flushing pressure	microhardness	TRD	The improvement in the surface finish was noted significantly.	[16]
Kumar et al.	Alumina powder blended EDM	Inconel 825; copper electrode		pulse on-time, current, gap voltage	SR, MRR, RLT	RSM	1. The average surface roughness obtained with powder blended EDM and conventional EDM was 2.8912 µm and 2.4169 µm, respectively. Furthermore, an average improvement of 0.4743 µm surface roughness was obtained by the proposed PMEDM. 2. At the centre values of parameters (peak current 5A, pulse on-time 7 µs, gap voltage 30V), the improvement claimed was 0.758 µm. 3. The improvement was recorded as a result of a larger inter-electrode gap due to the presence of powder particles, which facilitated better debris removal	[17]

Gr –Graphite; Al – Aluminium; HCHCr – High carbon high chromium; Si – Silicon; W – Tungsten; Cu – Copper; W-Cu – Tungsten copper; Br – Brass; RLT – Recast layer thickness; TRD – Taguchi's robust design; RSM– Response surface methodology

and copper. This research work also presents a comparison of surface roughness values obtained with the proposed method to those with conventional EDM for centre values of parameters along with analysis of machined surface through SEM micrographs.

3. MATERIALS AND METHODS

The experiments were performed on a Sparkonix EDM 35A machine. This traditional EDM machine was modified for electrode revolution. For this, a rotary die was fabricated and installed on the EDM head. This rotary die was integrated with the HMC AC servo motor through the chain and sprocket. After this, a separate tank of size (510 mm × 285 mm × 160 mm) was also fabricated to add silicon carbide powder of 60-micron particle size with EDM oil (kerosene). The powder concentration was fixed at 6g/l which yielded the finest surface during trial experiments. A small plastic impeller coupled with a stirrer motor shaft was used for continuously stirring the dielectric oil to ensure proper powder circulation in the tank. The stirrer rotational speed in a tank of 17 liters of oil (auxiliary tank) was measured using a digital tachometer (DT-2234B). The tachometer showed 565 rpm of the shaft. The machining of D3 steel was accomplished using a copper electrode. The size of the workpiece and electrode diameter were (47 mm × 31 mm × 7.5 mm) and 20 mm, respectively. Figure 1 illustrates the actual machining setup along with the fabricated rotary die and copper electrode. A comprehensive schematic diagram of the investigational setup is also presented in Figure 2. The elemental constitution of the work material is listed in Table 2. The randomization of run order was performed before conducting the experiments to distribute randomly the variance due to uncontrollable factors. The full factorial experimentation was done by taking three parameters, namely, peak current (Pc), pulse on-time (Pot), and electrode revolution (Erpm) to observe surface roughness as the output characteristic. The pulse-off time value was 30 µs during experiments. The total sparking time was kept fixed at 2 minutes for experimental runs. Table 3 displays the machining variables and their levels. The trials were done in triplicate before the surface roughness measurements were recorded. These values were also evaluated at three different sites on the workpiece's cavity. The average of these readings was noted as a

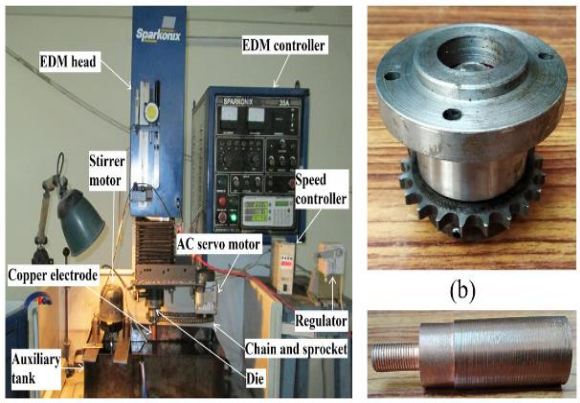


Figure 1. (a) Setup for experimental runs; (b) electrode revolution die; (c) copper electrode

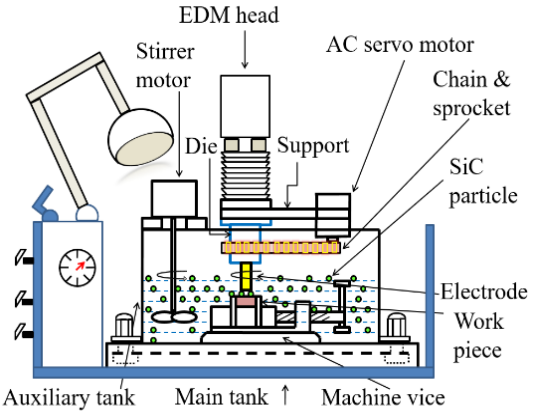


Figure 2. Schematic diagram of the investigational setup

TABLE 2. Elemental constitution of D3 steel

Cr	C	O	Si	Mn	S	P	W	V	Fe
13.40	2.23	2.0	0.41	0.40	0.03	0.05	0.15	0.08	81.25

TABLE 3. Machining variables

Parameter	Unit	Abbreviation	Level		
			1	2	3
Peak current	Ampere (A)	Pc	12	18	24
Pulse on-time	micro-seconds (μs)	Pot	100	400	1000
Electrode revolution	rpm	Erpm	1200	1500	1800

result of each experiment. For measuring SR, Tesa Rugosurf 10G tester was used by fixing a stylus traverse span of 0.8 mm each for 5 cut-offs. The results of surface roughness are presented in Table 4 along with their input factor level settings.

4. STATISTICAL MODELING OF THE RESPONSE CHARACTERISTIC

To set up a linkage between machining variables and SR of D3 steel, regression analysis of roughness data was done and consequently, a regression equation of 2nd order was developed. Analysis of variance (ANOVA) was done to discover the important terms and interactions between them along with their percentage contributions in surface roughness of D3 steel. The outcomes of ANOVA of SR after removing insignificant terms and interactions will be presented in the next section. Only terms conforming to a P-value (< 0.05) were considered as per 95% confidence interval [18]. During machining, noise due to uncontrollable factors affects the machining results. As a consequence, variation is likely to occur in data. Therefore, an assessment of the range for the predicted model was done encompassing the deviation due to noise. This deviation range [8] was computed by Equation (1). Therefore, prediction by the statistical model can be given by $Y \pm \Delta Y$, where ΔY is given as follows:

$$\Delta Y = t_{\alpha/2,DF} \times \sqrt{V_e} \quad (1)$$

where Y represents response characteristic, t represents the value of t-distribution corresponding to the specified significance level ($\alpha/2 = 0.05$, at 95% confidence interval) and degree of freedom of residual error (DF). V_e represents the residual error of the adjusted mean square. The surface roughness prediction model and the range equation for D3 steel are represented by Equations (2) and (3), respectively.

$$\begin{aligned} SR &= 12.976 + (0.2112 \times Pc) + (0.00489 \times \\ &Pot) - (0.006702 \times Erpm) - (0.000011 \times Pot \times \\ &Pot) + (0.000004 \times Pot \times Erpm) \end{aligned} \quad (2)$$

$$SR_{range} = SR \pm 0.92 \text{ microns} \quad (3)$$

Table 4 summarized 27 experimental runs with machining variables and outcome characteristics.

5. STATISTICAL MODEL VALIDATION

The validation of the statistical model was done by calculating the prediction value supplemented by range with the help of Equations (2) and (3) and thereafter, validating the values thus obtained with experimental results. The results of predicted and experimental findings for arbitrarily chosen run orders 20, 16, 11, and 25 are presented in Table 5. Furthermore, all other values

were also checked and found well within the range. Figure 3 shows the surface roughness versus run order plot for experiment and predicted values. The fitness of the regression predictor can also be ascertained from the distribution of residuals. Figure 4(a-d) depicts the surface roughness residual distribution diagrams illustrating normal probability chart, residual versus fitted value, vertical bar chart, and residual versus observation chart. From Figure 4(a, c), it is evident that all the residuals adhere to a straight line indicating normal distribution. The regression model is thus validated.

TABLE 4. Experimental runs

Run order	Machining variables			Output characteristics			
	Pc (A)	Pot (µs)	Erpm (rpm)	SR			Mean SR (µm)
				1	2	3	
1	12	1000	1500	5.73	6.41	5.59	5.91
2	18	100	1200	10.21	9.6	9.44	9.75
3	18	100	1500	8.86	7.35	7.82	8.01
4	12	100	1500	7.21	7.23	6.83	7.09
5	24	400	1800	9.64	9.89	9.93	9.82
6	12	400	1800	6.13	6.89	6.9	6.64
7	18	100	1800	5.6	6.2	5.42	5.74
8	18	400	1200	10.7	10.31	11.6	10.87
9	18	1000	1800	5.52	5.36	6.07	5.65
10	24	1000	1500	7.71	8.76	8.49	8.32
11	18	400	1500	9.87	11	10.09	10.32
12	24	100	1500	9.3	8.54	8.53	8.79
13	12	1000	1200	6.9	6.97	6.71	6.86
14	12	1000	1800	4.88	5.39	5.36	5.21
15	18	400	1800	7.63	8.71	7.51	7.95
16	12	100	1800	4.83	4.84	4.76	4.81
17	18	1000	1200	7.59	8.26	8.09	7.98
18	24	1000	1200	9.24	8.37	7.86	8.49
19	24	100	1800	6.19	6.7	6.43	6.44
20	12	100	1200	7.6	7.54	8.41	7.85
21	24	400	1500	11.46	10.86	11.04	11.12
22	24	400	1200	12.81	11.79	11.37	11.99
23	24	100	1200	11.36	11.74	11.13	11.41
24	12	400	1500	7.71	7.53	8.43	7.89
25	24	1000	1800	8.7	8.18	7.09	7.99
26	18	1000	1500	6.7	6.84	7.76	7.1
27	12	400	1200	9.11	9.42	9.37	9.3

TABLE 5. Statistical model validation

Run order	Machining variables			SR (microns)	
	Pc (A)	Pot (µs)	Erpm (rpm)	Predicted	Experiment
20	12	100	1200	8.38 ± 0.92	7.85
16	12	100	1800	4.54 ± 0.92	4.81
11	18	400	1500	9.54 ± 0.92	10.32
25	24	1000	1800	7.64 ± 0.92	7.99

6. RESULTS AND DISCUSSION

The effects of peak current (Pc), pulse on-time (Pot), and electrode revolution (Erpm) on SR of D3 steel along with their percentage share are portrayed in Figure 5. This was drawn by taking the mean value of SR at each level of Pc,

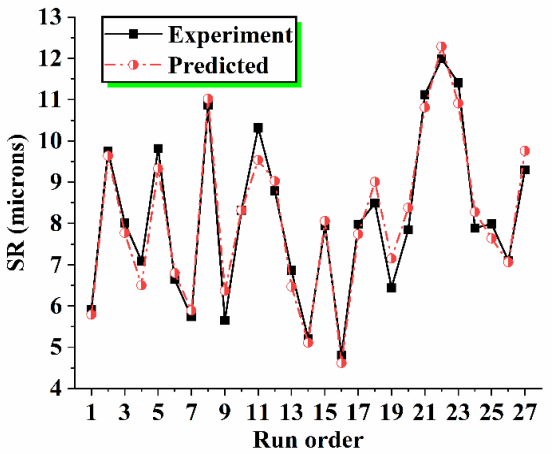


Figure 3. A comparative plot of surface roughness of experiment and predicted values

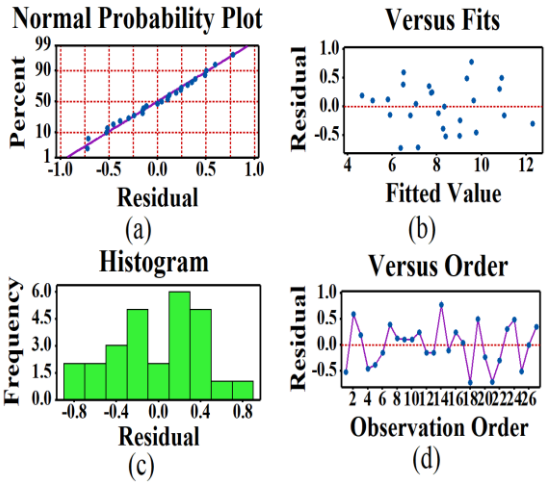


Figure 4. Residual plots of surface roughness (a) normal probability chart; (b) residual versus fitted value; (c) vertical bar chart; (d) residual versus observation order

Pot, and Erpm. Hence, three different mean values are shown for each variable. These mean response values are depicted in Table 6. Furthermore, Figure 6(a-c) depicts the plots of surface roughness versus electrode revolution at different peak currents of 12 A, 18A, and 24 A. From ANOVA outcomes in Table 7, it was found that the highest contribution was made by Erpm followed by Pc and Pot with 32.75, 28.98 and 6.09%, respectively.

6. 1. Effect of Peak Current Figure 5(a) illustrates the upshot of peak current on SR. It is observed that the rise in peak current results in an increase in SR. The reason may be assigned to the supply of higher energy at a high peak current which leads to deeper craters on the workpiece [19, 20]. These deep craters make the surface rougher, and hence, higher SR was observed [21].

6. 2. Effect of Pulse On-Time The effect of pulse on-time (Pot) on SR may be witnessed from Figure 5(a). The increase of Pot from 100 µs to 400 µs has been found to increase SR whereas, SR decreased with a further increase in Pot from 400 µs to 1000 µs. The reason may be credited to the widening of the plasma channel at a high Pot. This decreases the spark intensity, ultimately, leading to less energy impact on the work piece, and hence, less crater size [9, 22]. This smoothed the surface for large pulse on-time values such as 1000 µs.

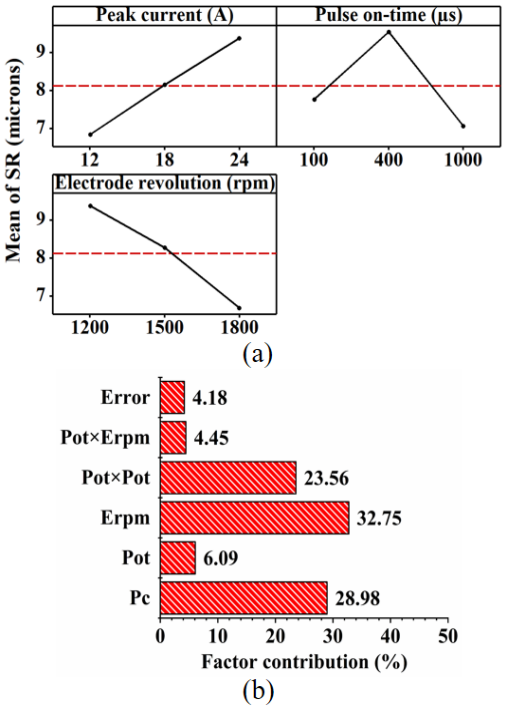


Figure 5. (a) Main effect plot and (b) percentage aid of machining variables on SR

TABLE 6. Response table of means

Level	Pc (A)	Pot (µs)	Erpm (rpm)
1	6.84	7.766	9.389
2	8.152	9.544	8.283
3	9.374	7.057	6.694
Delta	2.534	2.488	2.694
Rank	2	3	1

6. 3. Effect of Electrode Revolution It may be elucidated from Figure 5(a) and Figure 6(a-c) that the rise in electrode revolution resulted in improvement of rendered surface quality. The rationale of this achievement may be contributed to the efficient cleaning

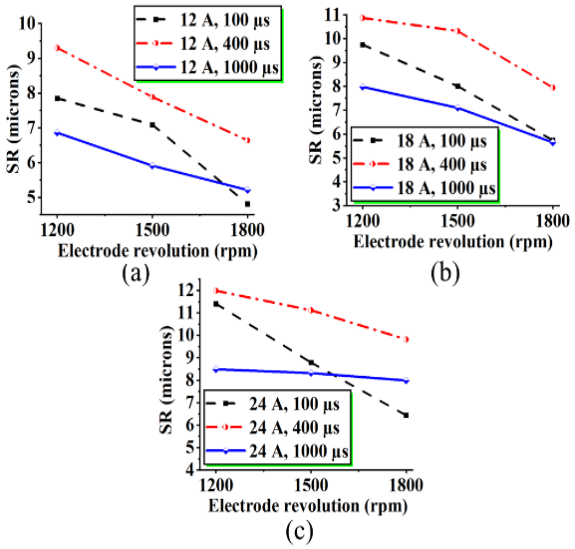


Figure 6. Effect of electrode revolution on surface roughness at peak current (a) 12 A; (b) 18 A; (c) 24 A

TABLE 7. Analysis of variance table for SR

Source	DF	Seq. SS	Share	Adj. SS	Adj. MS	F-Value	P-Value
Regression	5	95.57	95.82%	95.57	19.11	96.29	0.000
Pc	1	28.90	28.98%	28.90	28.90	145.61	0.000
Pot	1	6.07	6.09%	1.39	1.39	7.04	0.015
Erpm	1	32.67	32.75%	26.11	26.11	131.57	0.000
Pot × Pot	1	23.49	23.56%	23.49	23.49	118.36	0.000
Pot × Erpm	1	4.43	4.45%	4.43	4.43	22.34	0.000
Error	21	4.16	4.18%	4.16	0.19		
Total	26	99.74	100.00%				

DF: degree of freedom; Seq. SS: sequential sum of squares; Adj. SS: adjusted sum of squares; Adj. MS: adjusted mean square; F: Fisher's value.

of debris from the machining area. With the increase in electrode revolution, centrifugal force increases, and as the fluid also has abrasive silicon carbide powder, better abrasion, as well as fast cleaning, have benefited the surface, and hence, less surface roughness was obtained [13].

7. INTERACTION EFFECT

Analysis of variance has shown the interaction between pulse on-time and electrode revolution to be inevitable due to a lower P-value than 0.05. Figure 7 depicts the interaction and surface plot of pulse on-time and electrode revolution. The surface roughness of D3 steel was found to increase from 100 to 400 µs and then, decreased from 400 to 1000 µs for all values of electrode revolution. The lower electrode revolution has resulted in a rough surface than the higher rotational values. The minimum surface roughness was recorded at the lowest setting of pulse on-time (100 µs) and the highest setting of electrode revolution (1800 rpm), respectively. The rationale for this particular combination is once again reiterated to the better circulation of dielectric oil facilitated by higher centrifugal force at 1800 rpm of electrode revolution along with abrasive effects of the powder. Simultaneously, the surface finish was augmented by shallow craters due to low-intensity sparks at 100 µs of pulse on-time.

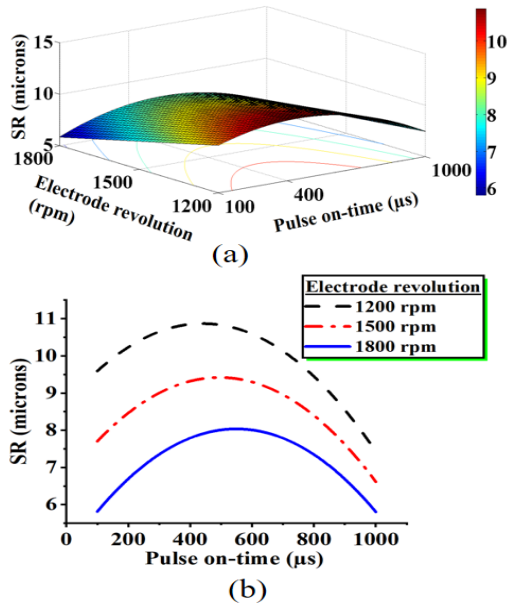


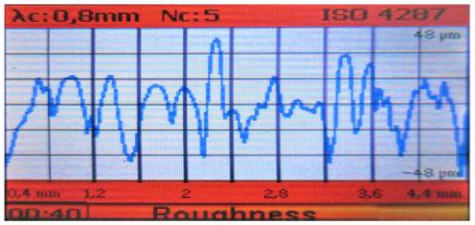
Figure 7. (a) Surface plot; (b) interaction plot of pulse on-time and electrode revolution on SR

8. SURFACE ROUGHNESS PARAMETRIC ANALYSIS

The parametric analysis of surface roughness profiles was executed at Pc-18A, Pot-400 µs and Erpm-1500 rpm with traditional EDM and SiC powder blended rotational EDM. Surface roughness profiles with details of parameters are depicted in Figure 8 for both processes. Three parameters total profile height (Rt), root mean square roughness (Rq), and roughness average (Ra) were measured. The reason for selecting these specific parameters was their capability to explain the entire profile at a glance and acceptance throughout the industries. These parameters in traditional EDM and SiC powder blended rotational EDM were found to be Rt-86.63 µm, Rq-15.85 µm, Ra-12.36 µm, and Rt-70.30 µm, Rq-12.45 µm, Ra-10.32 µm, respectively (refer Figure 8). With these results, the integration of SiC powder blending into EDM oil along with electrode revolution has outperformed the traditional method of EDM machining. The credit of this improvement goes to the efficient dispersal of debris from the inter-electrode gap and formation of shallow craters, facilitated by blending of powder and electrode revolution.

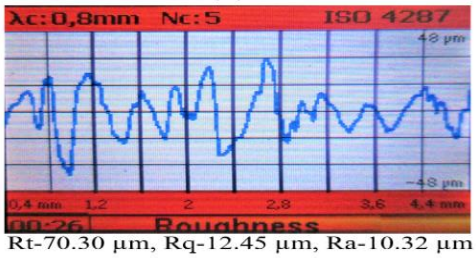
9. SURFACE MORPHOLOGY

Figure 9 shows the result of field emission scanning electron microscopy performed on machined surfaces by traditional EDM and SiC powder blended rotational EDM. The surface images were captured at a magnification of (500×) for both the processes and further, images at a magnification of (2000×) show the in-depth analysis of crater dimensions. The traditional EDM process resulted in a rough surface and had a lot of cracks and craters (refer Figure 9(a)). An addition of powder and electrode revolution have resulted in a fine surface with fewer cracks (refer Figure 9(b)). Furthermore, the in-depth crater dimension analysis revealed good results with silicon carbide powder blended rotational EDM. The maximum crater depth measured was 18.6 µm in traditional EDM whereas, it was 8.51 µm in SiC powder blended rotational EDM (refer Figure 9(c, d)). These results are a clear reflection of a better surface in SiC powder blended rotational EDM. The reason for the improved surface may be assigned to the efficacious expelling of debris by electrode rotation [13, 23] and higher gap owing to the existence of powder particles blended with EDM oil, which also facilitates effective flushing of debris[17, 24]. Along with this, the intensity of the sparks decreases because of a higher gap between the work piece and electrode leading to the shallow craters, and hence surface finish increases.



Rt-86.63 µm, Rq-15.85 µm, Ra-12.36 µm

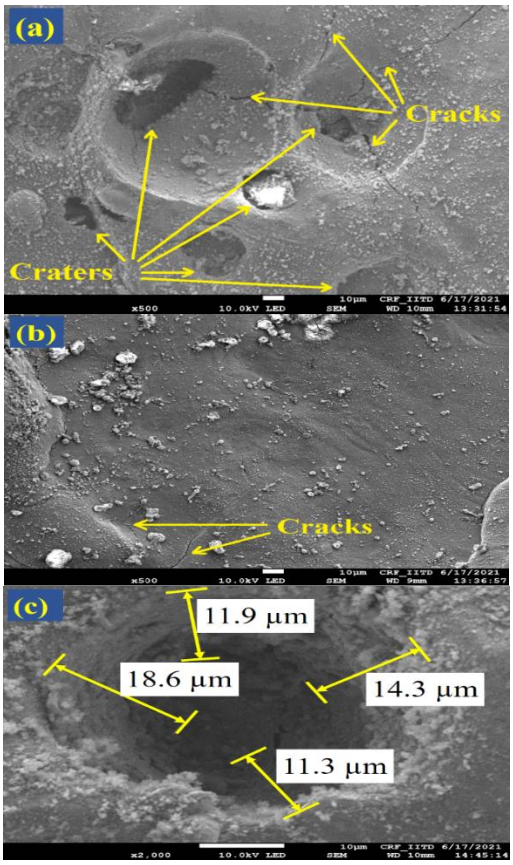
(a)



Rt-70.30 µm, Rq-12.45 µm, Ra-10.32 µm

(b)

Figure 8. Surface roughness profiles (a) traditional EDM; (b) SiC powder blended rotational EDM. (Pc-18A, Pot-400 µs, Erpm-1500 rpm). (Rt- total profile height, Rq- root mean square roughness, Ra- roughness average)



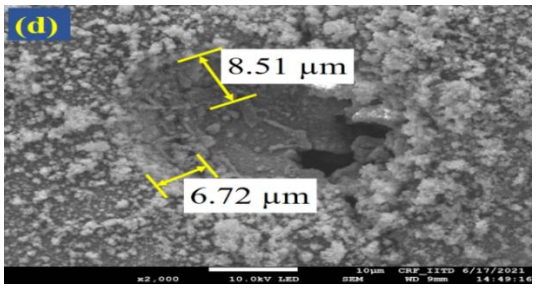


Figure 9. FESEM images of machined samples (a) traditional EDM (500×); (b) SiC powder blended rotational EDM (500×); (c) traditional EDM (2000×) (d) SiC powder blended rotational EDM (2000×). (peak current 18A, pulse on-time 400 µs, electrode revolution 1500 rpm)

10. CONCLUSIONS AND FUTURE SCOPE

The study on surface roughness of D3 steel with the copper electrode in SiC powder blended rotational EDM resulted in the following conclusions:

1. The surface roughness increased with peak current. An increase in SR with pulse on-time was also observed up to 400 µs and a decrease in thereafter.
2. SR reduced with an increase in electrode revolution. An improvement of 28.67% was achieved in the experimented range.
3. The statistical analysis to study the effect of EDM parameters on surface roughness was successfully done.
4. The parametric settings that favoured the highest surface quality were Pc-12A, Pot-100 µs, and Erpm-1800 rpm. At these optimal settings, surface roughness was 4.81 µm.
5. The surface morphology results justified the applicability of SiC powder blended rotational EDM in comparison to the traditional EDM as it led to the creation of fewer cracks and shallow craters on the surface of the work piece. The maximum crater depths were 8.51 µm and 18.6 µm in SiC powder blended rotational EDM and traditional EDM, respectively.
6. The surface roughness profiles recorded with SiC powder blended rotational EDM and traditional EDM have also shown an appreciable decrease in surface roughness (Ra) from 12.36 µm to 10.32 µm.

The future scope of work in the proposed area may be extended with analysis of inter-electrode gap. With exact estimation of gap distance, flushing of debris can be further improved. The spark travel can also be analysed under the influence of electrode rotation as its transverse and inclined travel might affect the discharge crater dimensions.

11. LIMITATIONS

The application of the proposed process is limited to only circular cavity either blind or through-hole as the electrode is under rotation while machining. So, it cannot be used for random cavity/impression generation like for die making.

12. DECLARATION

This research study was conducted without any funding from any agency.

13. REFERENCES

1. Ho, K. H., Newman, S. T. "State of the art electrical discharge machining (EDM)", ***International Journal of Machine Tools and Manufacture***, Vol. 43, No. 13, (2003), 1287-1300. DOI:10.1016/S0890-6955(03)00162-7
2. Qosim, N., Supriadi, S., Puspitasari, P., Kreshanti, P. "Mechanical surface treatments of Ti-6Al-4V Miniplate implant manufactured by electrical discharge machining", ***International Journal of Engineering, Transactions A: Basics***, Vol. 31, No. 7, (2018), 1103-1108. DOI:10.5829/ije.2018.31.07a.14
3. Puthumana, G., Joshi, S. S. "Investigations into performance of dry EDM using slotted electrodes", ***International Journal of Precision Engineering and Manufacturing***, Vol. 12, No. 6, (2011), 957-963. DOI:10.1007/s12541-011-0128-2
4. Kansal, H. K., Singh, S., Kumar, P. "Parametric optimization of powder mixed electrical discharge machining by response surface methodology", ***Journal of Materials Processing Technology***, Vol. 169, No. 3, (2005), 427-436. DOI:10.1016/j.jmatprotec.2005.03.028
5. Chattopadhyay, K. D., Verma, S., Satsangi, P. S., Sharma, P. C. "Development of empirical model for different process parameters during rotary electrical discharge machining of copper-steel (EN-8) system", ***Journal of Materials Processing Technology***, Vol. 209, No. 3, (2009), 1454-1465. DOI:10.1016/j.jmatprotec.2008.03.068
6. Tong, H., Li, Y., Wang, Y. "Experimental research on vibration assisted EDM of micro-structures with non-circular cross-section", ***Journal of Materials Processing Technology***, Vol. 208, No. 1-3, (2008), 289-298. DOI:10.1016/j.jmatprotec.2007.12.126
7. Abdullah, A., Shabgard, M. R. "Effect of ultrasonic vibration of tool on electrical discharge machining of cemented tungsten carbide (WC-Co)", ***International Journal of Advanced Manufacturing Technology***, Vol. 38, No. 11-12, (2008), 1137-1147. DOI:10.1007/s00170-007-1168-8
8. Srivastava, V., Pandey, P. M. "Experimental investigation on electrical discharge machining process with ultrasonic-assisted cryogenically cooled electrode", ***Proceedings of the Institution of Mechanical Engineers, Part B: Journal of Engineering Manufacture***, Vol. 227, No. 2, (2013), 301-314. DOI:10.1177/0954405412469487
9. Beravala, H., Pandey, P. M. "Experimental investigations to evaluate the effect of magnetic field on the performance of air and argon gas assisted EDM processes", ***Journal of Manufacturing Processes***, Vol. 34, (2018), 356-373. DOI:10.1016/j.jmapro.2018.06.026
10. Singh, G., Satsangi, P. S., Prajapati, D. R. "Effect of Rotating Magnetic Field and Ultrasonic Vibration on Micro-EDM Process", ***Arabian Journal for Science and Engineering***, Vol. 45, No. 2, (2020), 1059-1070. DOI:10.1007/s13369-019-04229-3
11. Shih, H. R., Shu, K. M. "A study of electrical discharge grinding using a rotary disk electrode", ***International Journal of Advanced Manufacturing Technology***, Vol. 38, No. 1-2, (2008), 59-67. DOI:10.1007/s00170-007-1068-y
12. Aliakbari, E., Baseri, H. "Optimization of machining parameters in rotary EDM process by using the Taguchi method", ***International Journal of Advanced Manufacturing Technology***, Vol. 62, No. 9-12, (2012), 1041-1053. DOI:10.1007/s00170-011-3862-9
13. Dwivedi, A. P., Choudhury, S. K. "Improvement in the Surface Integrity of AISI D3 Tool Steel Using Rotary Tool Electric Discharge Machining Process", ***Procedia Technology***, Vol. 23, (2016), 280-287. DOI:10.1016/j.protcy.2016.03.028
14. Bhattacharya, A., Batish, A., Singh, G., Singla, V. K. "Optimal parameter settings for rough and finish machining of die steels in powder-mixed EDM", ***International Journal of Advanced Manufacturing Technology***, Vol. 61, No. 5-8, (2012), 537-548. DOI:10.1007/s00170-011-3716-5
15. Bhattacharya, A., Batish, A., Kumar, N. "Surface characterization and material migration during surface modification of die steels with silicon, graphite and tungsten powder in EDM process", ***Journal of Mechanical Science and Technology***, Vol. 27, No. 1, (2013), 133-140. DOI:10.1007/s12206-012-0883-8
16. Singh, A. K., Kumar, S., Singh, V. P. "Effect of the addition of conductive powder in dielectric on the surface properties of superalloy Super Co 605 by EDM process", ***International Journal of Advanced Manufacturing Technology***, Vol. 77, No. 1-4, (2015), 99-106. DOI:10.1007/s00170-014-6433-z
17. Kumar, A., Mandal, A., Dixit, A. R., Das, A. K., Kumar, S., Ranjan, R. "Comparison in the performance of EDM and NPMEDM using Al 2 O 3 nanopowder as an impurity in DI water dielectric", ***International Journal of Advanced Manufacturing Technology***, Vol. 100, No. 5-8, (2019), 1327-1339. DOI:10.1007/s00170-018-3126-z
18. Moghaddam, M. A., Kolahan, F. "Improvement of Surface Finish when EDM AISI 2312 Hot Worked Steel using Taguchi Approach and Genetic Algorithm", ***International Journal of Engineering, Transactions C: Aspects***, Vol. 27, No. 3(C), (2014), 417-424. DOI:10.5829/idosi.ije.2014.27.03c.09
19. Prakash, K. S., Gopal, P. M., Rahul, R. N. "Effect of material and machining features in electric discharge machining of 6061Al/rock dust composites", ***Indian Journal of Engineering and Materials Sciences***, Vol. 27, No. 2, (2020), 471-480
20. Singh, S., Maheshwari, S., Pandey, P. C. "Some investigations into the electric discharge machining of hardened tool steel using different electrode materials", ***Journal of Materials Processing Technology***, Vol. 149, No. 1-3, (2004), 272-277. DOI:10.1016/j.jmatprotec.2003.11.046
21. Phan, N. H., Donga, P. V., Muthuramalingamb, T., Thiena, N. V., Dunga, H. T., Hunga, T. Q., Duca, N. V., Lya, N. T. "Experimental investigation of uncoated electrode and PVD AlCrNi coating on surface roughness in electrical discharge machining of Ti-6Al-4V", ***International Journal of Engineering, Transactions A: Basics***, Vol. 34, No. 4, (2021), 928-934. DOI:10.5829/ije.2021.34.04a.19
22. Beravala, H., Pandey, P. M. "Experimental investigations to evaluate the surface integrity in the magnetic field and air/gas-assisted EDM", ***Journal of the Brazilian Society of Mechanical Sciences and Engineering***, Vol. 43, No. 4, (2021), 213. DOI:10.1007/s40430-021-02929-2
23. Bajaj, R., Dixit, A. R., Tiwari, A. K., Chauhan, N. K. "Machining performance enhancement of en-31 diesteel using mwcnt mixed rotary edm", ***Indian Journal of Engineering and Materials Sciences***, Vol. 27, No. 2, (2020), 309-319.

24. Joshi, A. Y., Joshi, A. Y. "A systematic review on powder mixed electrical discharge machining", ***Heliyon***, Vol. 5, No. 12, (2019), e02963. DOI:10.1016/j.heliyon.2019.e02963

Persian Abstract

چکیده

این مطالعه تحقیقاتی بر روی اتحاد پودر کاربید سیلیکون به اندازه ٦٠ میکرومتر به روغن دی الکتریک (نفت سفید) با چرخش الکترود در ماشینکاری تخلیه الکتریکی فولاد D3 با استفاده از الکترود ساخته شده از مس متمرکز گردید. جریان پیک، زمان پالس و چرخش الکترود به عنوان متغیرهای ماشینکاری برای بررسی تأثیر آنها بر زبری سطح قطعه کار انتخاب شدند. آزمایش فاکتوریل کامل با ٢٧ اجرای آزمایشی انجام شد. نتایج تحلیل واریانس نشان داد که چرخش الکترود به‌عنوان تأثیرگذارترین متغیر با درصد کمک ٣٢/٧٥ درصد، پس از پیک جریان و پالس به‌موقع به‌ترتیب ٢٨/٩٨ و ٦/٠٩ درصد است. کاهش زبری سطح در سرعت چرخش الکترود بالاتر ثبت گردید. میکروسکوپ الکترونی روبشی گسیل میدانی نمونه های ماشین کاری شده انجام شد و نتایج بهبود قابل توجهی را در ویژگی های سطح نشان داد. پرداخت سطح از ١٢/٣٦ میکرومتر به ١٠/٣٢ میکرومتر در ماشینکاری تخلیه الکتریکی چرخشی مخلوط پودر کاربید سیلیکون نسبت به ماشینکاری تخلیه الکتریکی سنتی بهبود یافته است. حداقل زبری سطح به دست آمده ٤/٨١ میکرومتر در شرایط بهینه بود.

Peer-to-Peer Networking and Applications
https://doi.org/10.1007/s12083-022-01325-4

Systematic study of topology control methods and routing techniques in wireless sensor networks

Radhika Kavra[1] · Anjana Gupta[1] · Sangita Kansal[1]

Received: 14 August 2021 / Accepted: 10 April 2022
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2022

Abstract
Emergence of applications of wireless sensor networks in various fields over time requires to acknowledge existing solutions which have been proposed in history to overcome various challenges/ issues related to topology formation and routing of sensing data among the sensors. Sometime data transmission requires topology establishment or its awareness and sometime it does not. In order to obtain optimal routing topology or an optimal route according to some desired aim, a wide range of topology control methods and routing techniques/ protocols come into sight. This article is the very first attempt in offering the collective and combined review on unexplored topology control methods and routing techniques, year-wise. Presenting a comprehensive study of various energy efficient topology control methods involving the graph based methodologies, explicitly designed interference models and related algorithms which have not been covered in previous year survey papers and the detail review on most efficient network structure based classical hierarchical routing protocols, topology aware particle swarm optimization and ant colony optimization based routing techniques in both static and mobile wireless sensor networks. However, the mentioned techniques don't cooperate delay sensitive routing or timely data delivery, for this purpose, we additionally provide a detail review on delay-sensitive real-time routing protocols in this survey paper. Analysis of each technique has been put forth year-wise by mentioning contribution made by the proposed work, detail of methodology opted, merits and demerits. A brief comparative study among the existing techniques, discussion about the research gaps and persisting issues along with direction of future work have also been provided.

Keywords Wireless sensor networks (WSNs) · Topology · Graph algorithms · Optimization techniques · α–approximate solution · Time complexity

1 Introduction

Wireless Sensor Networks (WSNs) can be defined as an infrastructure-less networks, made up of few tens to thousands of battery-powered omnidirectional wireless sensors. These wireless sensors work together to monitor a region to obtain data about the environment and cooperatively pass their data through the network to a main location (sink or base station) where the data can be observed and analysed. A sink or base station acts like an interface between users and the network. One can retrieve required information from the network by injecting queries and gathering results from the sink [1]. WSNs have gained worldwide attention in recent years because of the facilitation of low-cost multi-functional wireless sensors that are small in size and communicate untethered using radio signals. Each sensor node consists of three subsystems: sensing, processing and transmission (see Fig. 1 redrawn from [2]). Each sensing subsystem has transducers in it to collect information about the environment. A processing subsystem has an embedded processor to process local sensed information and transmission subsystem has a transceiver embedded in it to exchange sensed/ processed information with external systems (base station). Some sensor nodes have two additional units for location finding and mobility so that they can carry out the tasks assigned to them [3]. So this way, sensors can be static or mobile depending on the application goals. These tiny sensor nodes, which consist of sensing, data processing and communication components, leverage the idea of WSNs based on collaborative effort of large number of sensor nodes.

Now-a-days wireless sensors are greatly engaged in tracking and monitoring tasks like indoor/ outdoor environmental

✉ Anjana Gupta
anjanagupta@dce.ac.in

[1] Department of Applied Mathematics, Delhi Technological University, Bawana Road, Rohini, Delhi, India

Published online: 10 May 2022

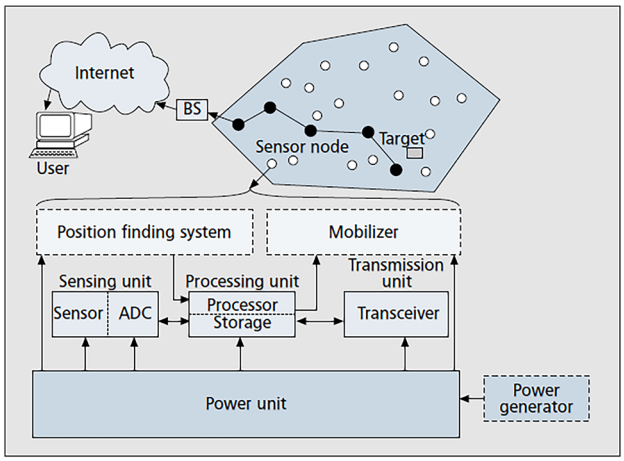
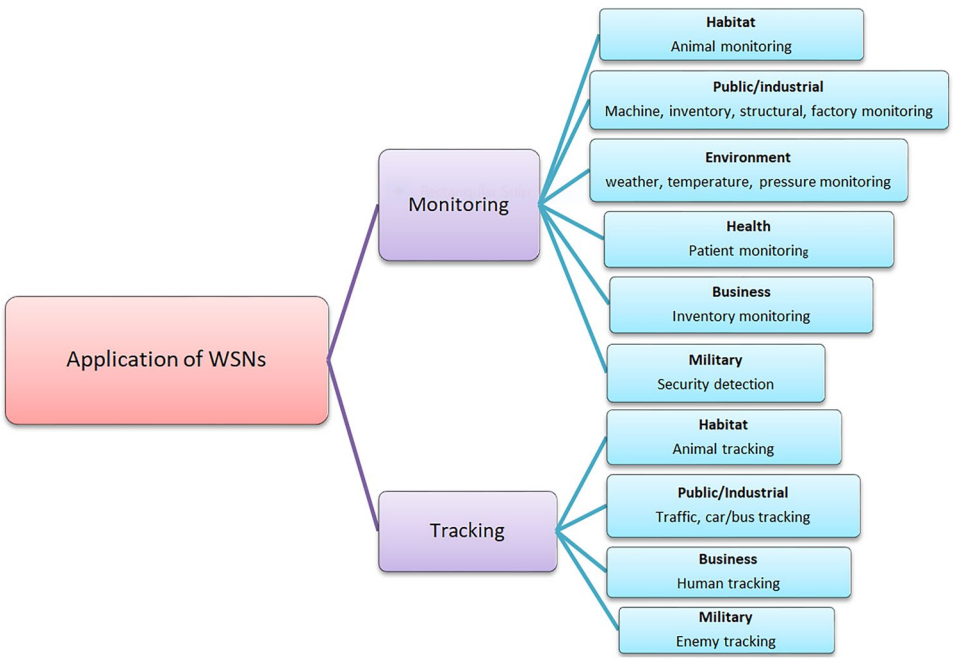


Fig. 1 Components of wireless sensors

monitoring, health and well-ness monitoring, power monitoring, inventory location monitoring, military surveillance, tracking objects, animals, humans and vehicles etc [4]. Figure 2 depicts the tracking and monitoring based applications of wireless sensors.

When sensors are deployed in a region, they are responsible for self-organizing an appropriate network infrastructure, called topology and often start multi-hop communication among themselves, called routing. Since each individual node in WSN is inherently resource constrained: it has limited processing speed, limited energy, communication bandwidth and storage capacity. Therefore more is the power expenditure of a node earlier it will die in the network since the only source of power for wireless sensors is battery, which is usually not re-chargeable or replaceable especially when the sensor nodes are expected to operate with no human intervention for a longer period of time. So engagement of wireless sensors in various application fields retain lots of limitations and challenges related to topology as well as routing in WSNs like connectivity, coverage, network lifetime, energy consumption, fault-tolerance, transmission delay and many more [5, 6], refer Fig. 3. Some of the issues like connectivity, coverage, node failure, energy consumption, scalability, fault-tolerance etc can be dealt at topology construction phase or underlying network acknowledgment phase of WSNs, these challenges are known as topology control (TC) issues and the issues which are occurred and tackled only at routing phase like transmission delay, transmission cost, reliability, throughput, load balancing etc, are called routing issues. There can be two situations regarding topology control and routing issues, either both can be dealt simultaneously or individually. As designing an energy-aware topology of randomly deployed wireless sensors results low cost routing of data packets but meanwhile it can cause higher transmission delay [7, 8]. Similarly designing a fault-tolerant topology increases route reliability with the increment in transmission cost. Therefore one can see, integration of topology control and routing solution can be helpful only if the proposed method hits the same target or fulfilling the desired goals. Routing in WSNs are done in both ways either using topology control or non-topology control depending upon the problem objective or the desired goal, for example cluster-based or hierarchical routing protocols involve topology awareness/ establishment

Fig. 2 Applications of WSNs



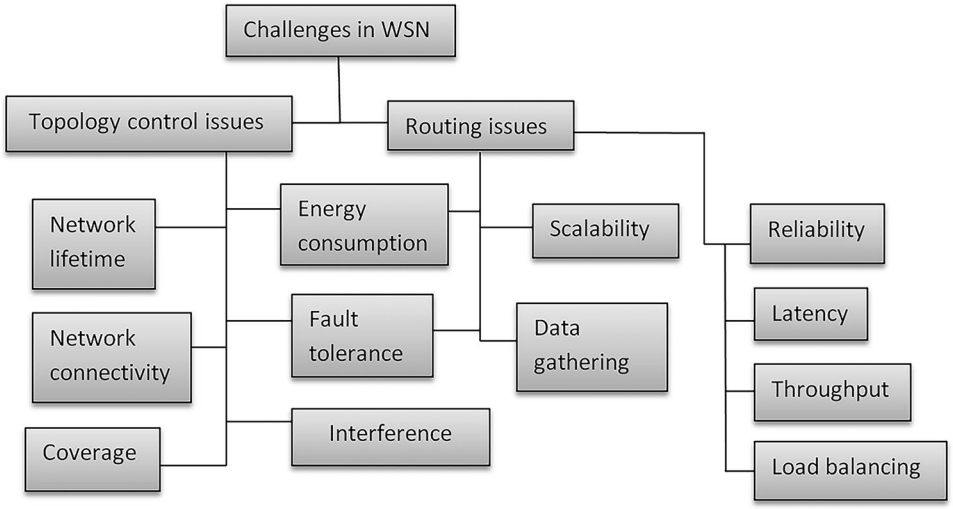


Fig. 3 Topology control and routing challenges in WSNs

to achieve energy efficiency [9, 10] while other routing techniques don't require topology setup [11, 12]. However, both topology control methods and routing techniques are used to enhance WSNs performance by saving energy of battery constraint sensor nodes in their own way, former works on topology and later is engaged to find end-to-end routing path between source to sink nodes.

1.1 Topology control and routing challenges in WSNs

Energy consumption: If the route from source sensor node to sink node is not an appropriate one for data transmission, more energy is consumed in the network. The energy required to transmit a $t-bit$ message over a distance d is given as follows:

$$E_T(t,d) = \begin{cases} t * E_{elec} + t * \varepsilon_{fs} * d^2 & for\ d < d_0 \\ t * E_{elec} + t * \varepsilon_{mp} * d^4 & for\ d \geq d_0 \end{cases} \quad (1)$$

where $E_{elec}, \varepsilon_{fs}, \varepsilon_{mp}$ be the energy required by the electronics circuit, amplifier and multi path respectively. Energy required to receive $t-bit$ message is $E_R(t) = t * E_{elec}$. Hence as data transmission takes place between the sensors, sensors start depleting their energy according to the Eq. 1.

Network lifetime: It is measured by the lifetime of sensor who dies first in the network due to early power drainage. So objective is to maximize minimum sensor's lifetime.

Network life cycle: Lifetime of lastly alive nodes from the beginning of network execution.

Network connectivity: Connectivity refers to the wireless approach of each sensor to atleast one other sensor using their transmission range to maintain the network connectivity. Loss of connectivity causes loss of sensors and/or forwarding data. So requirement of connected topology at every round of data transmission leads to energy expenditure in constructing and monitoring topology.

Interference: It is a kind of hindrance or obstacle existing in the topology due to collision of data packets transmitting over the same route and hence retransmission of data packets is needed. This causes more energy consumption of sensor nodes.

Load balancing: Some of the nodes are loaded more than others leading to fast energy depletion hence nodes get failed, loosing network connectivity and network lifetime. So load balancing refers to uniform distribution of sensing and transmission loads between the nodes of every cluster for a uniform power consumption to take place.

Coverage: Coverage refers to the area covered by a sensor to sense and process the information. A network needs to have an optimal coverage i.e. minimum number of sensors are deployed to cover whole target region. Lack of coverage in any part of target region causes energy hole in the topology i.e. no information is being collected and forwarded from that part of region due to loss of coverage.

Data gathering: It must be done in such a way that reduces data redundancy and data retransmission so that nodes' energy could be conserved.

Energy hole or Hot-spot: Nodes near sink, have to spend extra power in data aggregation and/or transmission. So they die very early and their space leave energy hole or hot-spot in the network with degrading routing process.

Latency: It refers to an end-to end delay in data delivery. Real-time applications of WSNs require least delay in data transmission so latency needs to minimize there.

Scalability: For applications where tens to thousands of sensor nodes are deployed in a target area, routing proto-

cols should be designed in a way that they will be able to handle and respond to such greater number of events or addition or deletion of nodes.

Reliability: Reliability of a WSN is related with the probability of successful end-to-end communication i.e. data is received by the sink successfully.

Throughput: Throughput refers to how much data can be transferred from source to destination within a given time frame.

Packet delivery ratio: It is the percent of data packets delivered to the base station to those generated by source nodes.

Quality of service (QoS): QoS is related to reliability, packet delivery ratio and throughput factor with minimum error in transmitted data.

Deadline miss ratio (DMR): It refers to the fraction of data packets which don't reach destination node within a given time-frame.

Fault tolerance: It is the ability of routing protocols to respond quickly and discover new routes for the transmission of data in case there are nodes failures in the network.

The root cause of all the challenges mentioned here is limited power storage of each sensor node. Moreover, WSNs which aim to minimize latency, maximize throughput or QoS, require more nodes' power to achieve the goals. Hence this brings out the tradeoff between energy and delay in real-time application based routing. As every task from topology setup to data transmission requires nodes energy that is why topology control and routing issues become challenging to handle in WSNs.

1.2 Optimization in WSNs

Limited available energy to sensor nodes has attracted the attention of many researchers to deal with energy consumption of sensor nodes, network lifetime, network connectivity, link interference, data transmission cost, data reliability, data latency, sensing coverage, load balancing, data gathering, fault tolerance in designing the network topology and routing protocols. These pertaining issues in WSNs give rise to many optimization problems whose objectives can be (any combination of) maximize energy efficiency, maximize link quality, maximize network lifetime, maximize reliability, maximize coverage, maximize throughput, minimize delay, minimize cost, minimize interference, minimize packet error rate, minimize unbalanced energy consumption etc subject to the constraints (any combination of) network connectivity, interference, quality of service (QoS), transmit power, distance, coverage, (static/mobile) topology etc. Therefore, optimization in WSNs are of 'Network' and 'Route' optimization kinds. Network optimization is related to achieve a desired topology goals of randomly distributed sensors in a target area that minimizes energy consumption or maximizes network lifetime subject to the constraint of network connectivity and/or coverage. Route optimization aims to find an optimal routing which results in less energy consumption fulfilling application based WSNs requirement like minimum latency, maximum throughput etc. In this respect, various topology control algorithms [13–18] targeting network optimization and routing protocols [5, 6, 19–21] targeting route optimization have been proposed in the literature for many past years. Existing literature reveals that energy conservation has been a major goal in designing topology and routing schemes in WSNs. There exists many survey papers/ review articles in history providing the study of topology control methods to optimize WSNs concerned with different objectives. Similarly, survey papers on routing protocols/ techniques with different optimization objectives and classifications are existing good in number from many past years. Among the existing survey papers/ review articles, one can either go through the study of topology control methods or the study of routing techniques at the same moment, i.e. no survey paper has provided the combined and collective study of both topology control methods and routing techniques. This motivates us to come-up with a review article offering the combined study of various energy efficient topology control methods and routing techniques along with the detail description of methodology opted, their merits and demerits along with future scope.

1.3 Contribution

Here, we are listing the contribution and focus of this review article. The major contributions of this paper are as follow:

1. This paper come-up with the combined and collective study of topology control methods and routing techniques in WSNs characterising the parameters of energy efficiency, transmission cost, nodes power consumption, network lifetime, network connectivity, coverage, load balancing, latency, QoS, scalability, fault-tolerance, solution complexity, mobility and immobility of network.
2. A taxonomy of topology control methods and routing techniques has been proposed with the aim to optimize the performance of WSNs both at topology level and routing level.
3. The basic idea to provide the thorough study of any classified category is its evolutionary techniques, well establishment and applicability in real-life scenario.
4. Year-wise representation of each reviewed technique/ protocol has been followed in this review article with the thorough mention of contribution made by the proposed work involving the beneficial part, specific approach/ methodology opted and its limitations which might be overcome in future work.

5. Study and analysis of not previously reviewed graph based approaches, explicitly designed interference model and related algorithms, is our prime concern towards topology control in WSNs.
6. Furthermore, insights on state-of-the-art network structure based and topology aware routing protocols; namely classical hierarchical routing, particle swarm optimization (PSO) and ant colony optimization (ACO) based routing, have been provided.
7. Review on real-time communication based routing protocols has been thoroughly presented, covering the articles from year 2002-2021, dealing with delay constraint, time bound, throughput, deadline miss ratio and reliability.
8. A brief comparative study among the proposed solutions has also been provided with the discussion on research gaps and open issues with new research directions for further development in this domain.

The rest of the article is organized as follows, Sect. 2 presents preliminaries, Sect. 3 introduces topology control methods and routing techniques in WSNs following the proposed classification, Sect. 4 offers the research methodology to conduct literature survey, Sect. 5 provides the summary of related survey papers from past years, Sects. 6 and 7 presents the review of various topology control methods and routing techniques respectively. Discussion part and comparative study can be seen in Sect. 8. Section 9 concludes the paper.

2 Preliminaries

Before going further in this review article, we would like to introduce general terminologies, graph related models, required definitions and abbreviations.

Topology control (TC) Topology control in WSNs is the problem of determining an appropriate topology or network structure with different topological goals like minimum interference and energy consumption with increment in network capacity like network lifetime, energy efficiency, scalability while maintaining the network connectivity and/ or coverage. TC is the initial stage for any topology aware routing protocol.

Routing in WSN Data transmission is called Routing. Routing can be done either using **single-hop** i.e. direct data transmission from source to sink or **multi-hop** i.e. data is transmitted node by node to reach sink.

Routing protocols Methodology/ set of rules that has been opted to find an optimal (best possible) route/ path from a source sensor node to the sink node. **Route optimality** may vary with objective of the problems concerned with different challenges like energy consumption, latency, reliability, throughput etc.

Centralized method A methodology which requires a master node to run and control other nodes in the network. Every other node needs to take permission or informs to master node first to work further.

Distributed method In this, each node has same authority to work so methodology can be initiated by any node independently as per the requirement.

Homogenous and Heterogenous WSN A WSN is said to be homogeneous if all the sensors have same storage, processing, battery power, sensing, and communication capabilities while a WSN which is not homogenous is said to be heterogenous.

Static and Dynamic/ mobile WSN A WSN where each sensor node even sink node is not allowed to move or kept fixed after deployment is called static WSN while in a dynamic or mobile WSN, nodes can move and if nodes are static then atleast sink node is mobile in nature.

Graphical layout of WSNs $G(V, E)$ WSN can be represented as directed/ undirected graph $G(V, E)$, with each sensor is taken as vertex (node) forming a vertex set V and communication link between sensors is taken as an edge in the plane forming an edge set E.

Weighted Graph modelling Weight on an edge can be defined as data transmission cost, link interference, Euclidean distance or delay in data transmission etc depending upon the objective of the problem which is to be solved.

Transmission cost function c A cost function c is defined as $c : V \times V \rightarrow R^{+} \cup \{\infty\}$ such that $c(u, v)$ is the power emission necessary for node u to send packet to node v in directed graph where $u, v \in V$. In undirected/ bidirected graph, $c(u, v) = c(v, u) = max\{c(u, v), c(v, u)\}$ can be taken symmetrically [15] and written as $c(uv)$ where $uv \in E$. $c(u, v) = \infty$ implies node u cannot communicate to node v directly.

Power dissipation over a link (uv) Power required for a direct transmission from u to v is a polynomial function of the distance d between u and v i.e. $c_1 + c_2 d^{\alpha}$, $where\ d = ||uv||$, α is the path loss gradient depending on transmission environment, c_1 and c_2 are some constant depending on electronic characteristics of wireless devices [22].

Power function *r* A power function $r : V \rightarrow R^{+}$ is used to assign sufficiently high transmission power to a node u so that it can send a message to node v directly i.e. $r(u) = max_{uv \in E} c(u, v)$ where E is the set of all edges [13].

Transmit power management based TC Establishing and controlling the topology of underlying network of WSNs by minimizing the nodes power consumption either by reducing redundant transmission links or managing cost efficient transmissions depending on the required aim. This consequently results reduced and appropriate transmit power to each node in the resultant topology. Details have been discussed in Sect. 6.1 and many related approaches have been documented in Table 1.

Transmission disk $D(u, r_u)$ of node *u* A circular disk $D(u, r_u)$ is assumed to exist, taking node u as a centre and $r_u = r(u) = max_{uv \in E} c(u, v)$ as a radius. Every other node which comes inside this disk is known as direct-neighbour of node u.

Coverage of a link uv and of a node u via link uv Coverage of link uv is $cov(uv) = \{w | w$ is covered by $D(u, r_{uv})$ or $D(v, r_{vu})\}$ and coverage of a node u via link uv is $cov_{uv}(u) = \{w | w$ is covered by $D(u, r_{uv})\}$. Coverage illustration is given Fig. 4, redrawn from [15].

Coverage based Interference of a link uv Interference of a link uv is denoted by $IC(uv) = |cov(uv)|$. According to the illustration given in Fig. 4, $IC(uv)$ comes out to be 11.

Node interference via link The interference of a node u is the maximum interference of all links incident on u i.e. $IC(u) = max_{uv \in E} IC(uv)$.

Sender-centric and receiver-centric interference Sender-centric interference happens to exist at the sender node u and it is denoted by $I_S(u) = \{v \in V \setminus \{u\} | v \in D(u, r_u)\}$ while the receiver-centric interference exists at receiver node v and it is denoted by $I_R(v) = \{u \in V \setminus \{v\} | u \in D(u, r_u)\}$.

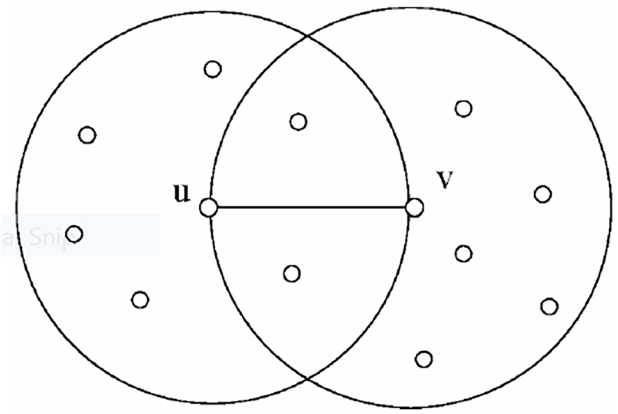


Fig. 4 Coverage of a link uv

Description of abbreviations			
BS	Base Station	RP	Rendezvous point
r_{min}	minimum transmission range	CH	Cluster head
r_{max}	maximum transmission range	MS	Mobile sink
SN	Sensor node	SP	Sojourn point
MN	Member nodes	ngh	neighbourhood

3 Classification of Topology control and routing techniques in WSNs

A proposed taxonomy of topology control and routing techniques is demonstrated in Fig. 5. Starting with the classification of topology control methods, Graph based approaches include those methods/ algorithms which are either depending on some graph-based model or utilizing existing concept and/or algorithms of graph theory like the concept of minimum spanning tree (MST), hamiltonian graph/ cycle, shortest path methods etc. For example, Yuanyuan et al. in [23] utilize the concept of dominating set of graph theory to obtain a connected topology, Abu-Affash et al. in [24] utilize the existence of second hamiltonian cycle to obtain low-powered connected network, Abderrahim et al. in [25] proposed Dijkstra algorithm based energy efficient technique, Yang and Cai in [26] proposed minimum spanning tree (MST) based QoS required topology likewise Cheng et al. in [27] proposed MST based algorithm to achieve Strong Minimum Energy Topology (SMET). Some algorithms are proposed to construct Relative Neighbor Graph (RNG) [28], Gabriel Graph (GG) [29], Yao Graph or Voronai diagrams to have an energy efficient spanner topology.

Additionally, Power control algorithms include different data gathering mechanism using transmit power control and/or power mode approaches to achieve energy efficient topology. Transmit power control approaches involve the control and adjustment of nodes' power level to achieve an appropriate transmission range so that nodes' energy can be conserved. Power mode approaches involve sleep/wakeup schedule of nodes means nodes can go to sleep/inactive mode if their tasks are fulfilled for a timing or their energy level is about to go zero or below some defined threshold. Some of power control algorithms with the engagement of sink nodes are MLS [30], PCLT [31], A3 [32] and some without sink nodes are DFTM [33], DPV [34], GBP [35]. Furthermore, Interference is one of the major issue in

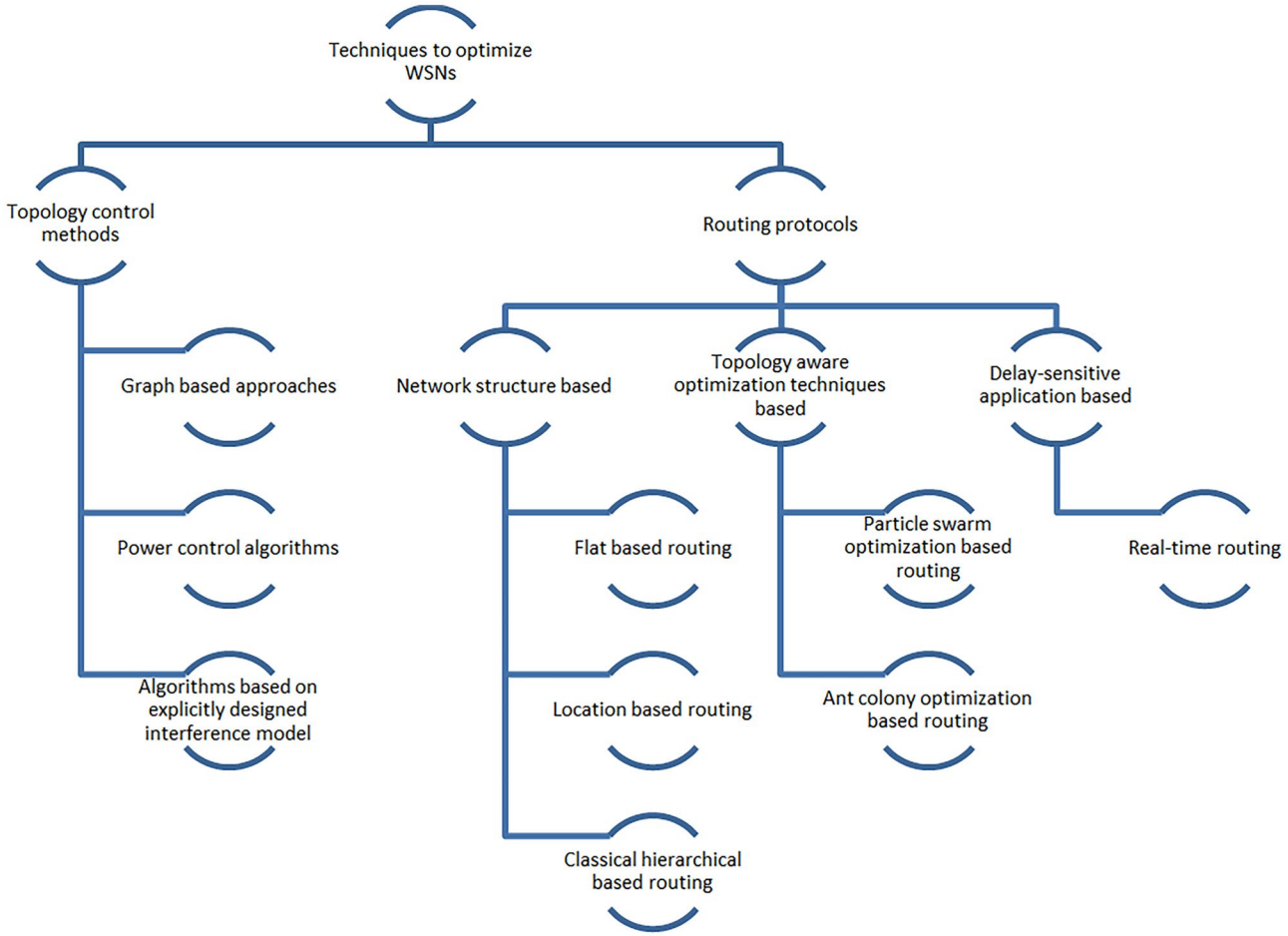


Fig. 5 Classification of Topology control and routing techniques to optimize WSNs

designing an energy efficient topology. Therefore various topology control algorithms minimizing interference explicitly based on interference models have been proposed in literature [15, 17, 22, 36–38]. Detail review on graph based approaches and algorithms to minimize transmit power and interference explicitly which have not been covered in the survey papers from past years, are presented in Sect. 6. Moving to the classification of routing protocols. On the basis of topology/ network structure formation, routing in WSNs has been classified into flat-based, location-based and hierarchical-based. Flat-based routing protocols have been proposed for homogenous WSNs where each sensor node has same functionality and operational capacity. It is also known as data-centric routing because base station sends queries to certain regions and waits for data from the sensors located in the selected regions. Some popular data-centric or flat-based routing protocols are Flooding, Gossiping [39], SPIN [40], Directed Diffusion [41], Rumor routing [42], Gradient based routing [43], COUGAR [44], ACQUIRE [45], MCFA [46]. However, flat-based routings show the limitations regarding reliable data delivery and fault-tolerance, result routing overhead and latency with the increase in node density. In location-based routing, also called geographic routing, location information of sensor nodes are required for route discovery, data forwarding and maintenance of sensor nodes in network. Energy consumption can be estimated by calculating distance between two particular nodes that helps in reducing energy consumption and optimizing whole network. GEAR [47], GAF [48], MECN, SMECN [49], EAGRP [50], LBRP [51] are some energy aware location based routing protocols. Existing strategies tend to increase extra routing overhead in order to obtain location information of destination nodes through the location service algorithm especially in dynamic WSN. In addition, energy hole problem and location inaccuracy problem also occur in geographic routing [6]. To tackle these issues, a novel fuzzy logic-based geographic routing protocol (FLGR) [52] has been proposed. Compared with existing protocols, the FLGR protocol has reduced routing overhead and increased packet delivery ratio in a sparse network. Location-based routings are having drawbacks in large dynamic WSNs since requirement of accurate address of each sensor node makes the

routing topology costly. Also performance degrades in term of latency in large network Hierarchical routing protocols also known as cluster-based routing, operate by clustering nodes in groups, assigning a cluster head (head of a cluster group) to gather data from the member nodes (remaining nodes other than CHs) in a cluster and forward it to the destination node. This causes reduction of redundant data transmissions which minimizes congestion at the destination node and energy consumption. Classical hierarchical routing protocols are mainly operated in two phases, first one is of cluster heads (CHs) selection and cluster formation phase either as chain or tree or grid structure. The second phase is of data aggregation and data transmission [20, 53]. In cluster-based routing, data transmission happens at two levels, first is called intra-cluster routing and second is inter-cluster routing. Intra-cluster routing is concerned with data transmission within a cluster among the member nodes (MNs) and inter-cluster routing is concerned with data transmission among the distinct clusters to sink nodes either through CHs or relaying nodes. Classical hierarchical routing protocols opted different strategy of CHs selection, cluster formation, data aggregation and data routing, aiming energy efficiency, scalability, fault-tolerance, load balancing, reliability and latency. These routing protocols outperform flat-based and location-based routing protocols in saving energy, reliability, load balancing, latency, scalability and extending lifetime of mobile WSNs (MWSNs). Due to more application and advantages of cluster-based routing, we limit our review on classical hierarchical routing protocols in this article from the classification of network structure based routing protocols. Optimization techniques based Particle Swarm optimization (PSO) and Ant Colony optimization (ACO) routing techniques are actually nature-inspired routing algorithms [54]. These techniques mimic the intelligence and social behavior of natured species like swarm and ants as they develop knowledge on their own to communicate and feed themselves. These routing algorithms resolve complex clustering and routing issues very efficiently in large dynamic WSNs [55]. As these techniques are robust, adaptable and flexible with respect to topology changes so they are called topology aware optimization techniques. In this survey article, review and analysis of latest routing algorithms based on PSO and ACO techniques is presented in Sect. 7. Irrespective of most energy efficient, scalable, fault-tolerant and robust, the mentioned techniques lag where the topology and routing requirement are delay-sensitive, bound with time, concerned with data delivery ratio, throughput and quality of service (QoS). Therefore, real-time routing techniques have been proposed to satisfy QoS requirements in term of minimum latency and maximum throughput. The main concern of real-time routing protocols is to meet packet delivery deadline as well as to maximize total throughput. Therefore real-time routing aims to discover an optimal route from source to destination meeting the real time constraints such as delay and throughput. Research in this field is on its early stage and limited to some well-known protocols like RAP [11], SPEED [12], MMSPEED [56], RPAR [57], THVR [58], EE-ARP [59], ERTLD [60], DCEMRA [61]. Emerging application of WSNs requires the real-time communication based routing protocols on priority so this survey provides the readers a whole knowledge and analysis of these protocols in Sect. 7.

4 Research methodology for literature review

Literature review on topology control methods and routing techniques in WSNs has been conducted to present articles that are published in well-established journals and conferences providing essential insights to researchers and practitioners working in this area. In our manuscript, we tried to summarize contributions, results, merits, demerits and future extension of the published work whose methodologies/ approaches are related to graph and optimization in WSNs. For the sake of understanding, flowchart of research methodology is demonstrated in Fig. 6. We searched web of science to obtain an extensive bibliography of the vast literature related to topology control methods and routing techniques in WSNs. The web of science core collection information provides us with fast and impactful access to the world's leading citation database such as ScienceDirect, IEEEXplore, Springer link online libraries, ACM digital library, Wiley, Research Gate and so forth.

The following subsections clearly outline the procedure we followed in extracting articles along with selection criteria, classification method and further explains the filtering processes.

4.1 Procedure for extraction of articles and selection criteria

We conducted a comprehensive search for articles using research questions, topic related terms and phrases such as 'Topology control in WSNs', 'Topology control algorithms/ techniques/methods in WSNs', 'Topology control and routing techniques in WSNs', 'Energy efficient hierarchical routing techniques/protocols in WSNs', 'Energy efficient and delay sensitive routing techniques', 'Graph based optimization techniques in WSNs', 'Real-time communication based routing protocols in WSNs', 'PSO and ACO techniques based routing in WSNs' and 'Topology and route optimization in WSNs'. The number of potential scholarly articles had been obtained at the initial step by using keywords described above was approximately 600, published in the year between 2000 and 2021. Many survey papers are

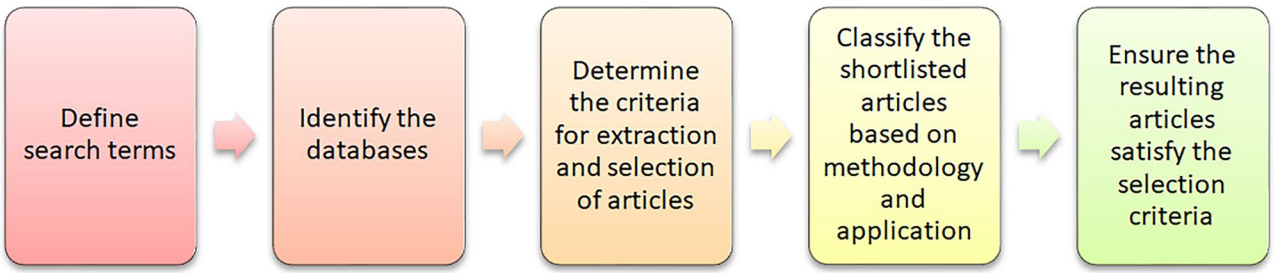


Fig. 6 Flowchart of research methodology

found from the obtained scholarly articles as well. So filtering of relevant survey papers are done first on the basis of the criteria given below.

- Only those survey papers are kept for review which have covered the literature of mentioned terms and phrases in a systematic and comprehensive manner.
- Also recent survey papers from the respected journals are kept by removing old survey papers covering the same topics.

After extracted relevant survey papers, we take note of already reviewed topology control techniques and routing protocols so that uncovered literature part can be acquainted and filtered out from the remaining scholarly articles for further review. Three criteria were determined and used to choose and acknowledge the respected journal and conference articles for additional consideration. Articles were excluded if they did not meet the accompanying selection criteria.

- From the obtained scholarly articles, only those articles are kept for review which are either containing topology control techniques or routing protocols including cluster-based/ hierarchical, real-time, PSO and ACO based routings in WSNs.
- Excluding techniques based on game-theory, fuzzy logic, machine learning, image processing related concept, entropy and statistics. Extracted articles have a basic foundation which is networking involving graph-theory based approaches, algorithmic approaches based on graph models and some optimization techniques.
- Filtered articles are crosschecked if they are already reviewed in one of the extracted survey papers in detail. Only those articles are carried forward which are not reviewed in details and left to review.

After examining the title and abstract of approximately 600 articles so obtained at the very initial step, we retrieved 145 articles to review thoroughly (including survey articles as well) which satisfies the selection criteria. These articles undergo the next round of classification based on methodology and application area. Then finally kept them for reviewing process.

5 Related work

There has been a lot of survey papers existing in literature on topology control methods and routing techniques in WSNs. Some of them are reviewed and documented below.

5.1 Previous survey on Topology control methods

Santi in 2005 [13] provides an overview to obtain power controlled topology of wireless sensors based on homogenous and non-homogenous range assignment. This paper has covered the literature concerned with building and maintaining connected topology of minimum energy cost. Description about static and mobile topology in term of graph model and range assignment along with proposed solutions and complexity has been provided. Routing graphs in 2-D which are Relative Neighborhood Graph (RNG), Gabriel Graph (GG), Delaunay Graph (DG), Yao Graph (YG), OrdYaoGG have been explained and analysed in term of optimal power spanner. This paper emphasises on bounded node degree to minimize power consumption i.e. lesser the maximum node degree more power can be conserved and no routing graph works good in this respect. Rickenbach et al. in 2009 [14] present the explicit study of interference reduction in both static and mobile WSNs by analysing and comparing the various proposed sender-centric and receiver-centric interference-based models and algorithmic approaches along with their properties and complexities. It has been stated that LIFE (Low Interference Forest Establisher) algorithm builds minimum interference connected topology, LLISE (Localized Low Interference Spanner Establisher) algorithm builds a minimum interference connected and spanner topology in localized manner, NCC (Nearest Component Connector) algorithm builds minimum interference tree topology with sink nodes. Authors put forth the complexity of finding interference bounds while maintaining connectivity on

developed 1-D interference models namely the exponential node chain and the generalized highway model. Aziz et al. in 2012 [62] provide a deep discussion of energy efficient and network lifetime maximizing distributed topology control techniques in 2-D and 3-D. The classification of techniques has been done into four categories based on power adjustment, power mode, clustering and hybrid following detail explaination of each mentioned techniques respectively. MECN, SMECN, COMPOW are power adjustment based techniques, GAF, STEM, ASCENT are power mode based techniques, PACDS, ECDS, TMPO are cluster based technique and SPAN, CLUSTERPOW, LEACH are hybrid techniques of power adjustment/ power mode and clustering. Authors offer performance analysis, merits and demerits of each mentioned technique, existing issues and future scope. Yetgin et al. in 2017 [63] come-up with a review of only lifetime maximization techniques in large-scale WSNs. In this survey paper, one can see many definitions of network lifetime based on different objectives with the board classification of network lifetime techniques into resource allocation, opportunistic transmission schemes, sleep-wake scheduling, routing/ clustering, mobile relays and sinks, coverage and connectivity, optimal deployment, data gathering/ network coding, data correlation, energy harvesting and beamforming based on the defined objectives. Authors mention the constraint functions, optimization tool and achieved optimality of the most popular techniques from year 2000 to 2015. Also Nguyen in 2020 [65] provides a review on network lifetime maximizing techniques in mobile WSNs. Mathematical definition of network lifetime has been demonstrated in different category. Classification of developed techniques has been based on nodes mobility, sink mobility and event mobility. The nodes mobility category is further classified based on strategies like coverage, connectivity, energy aware, cooperative computing and localized and clustering as well as sink mobility category has been classified into random, fixed and controlled mobility. Authors cover the well established techniques till 2019 by mentioning their objectives and methodologies along with possible issues, limitations and future scope. Mohamed et al. in 2018 [64] show the study of various models of WSNs, topological designs and routing protocols based on different real-life applications of WSNs. A detail taxonomy of WSN applications considering all the main design aspects and related network features has been presented by authors. Topological designs and routing protocols have been analyzed following the applications based classification like environment application, agricultural, military and health care application etc. Authors point out the basic methodology, special features and persisting issues of well-established routing protocols in term of energy cost, scalability and network lifetime. However, the survey is majorly focused on homogenous WSNs and its related routing protocols. Singla and Munjal in 2020 [18] present topology control algorithms by categorising into centralized and distributed algorithms based on opted data gathering mechanisms. Authors classify Minimum Load Set (MLS), Probability Distribution and Competition in Same Layer (PLCT), Secured Energy Conserving Slot based topology (SECS-TMP) as centralized algorithms. A3, Dead End Free Topology Maintanence Protocol (DFTM), Interference Based Topology Control Protocol (ITCD), Disjoint Path Vector (DPV), Grid Based Topology Control (GBP), Localized Topology Control (LDTC), Location Free Topology Control (LFTC), Hybrid AOA/TOA Topology Control protocol (HDTC), Cone Based Topology Control (CBTC), Articulation Point Topology Control (APTC), Light Weight Redundancy Aware Topology Control Protocol (LRTCP), Lossy Link Aware Opportunistic Topology Control (LOLA-OTC) come under distributed algorithms. Working mechanism and comparison of all the mentioned protocols have been provided. Authors involve various parameters to evaluate performance that are energy consumption, network lifetime, energy cost, delay, throughput, routing overhead, time, collision ratio, maximum relative load. Only few of them show good performance against every parameter. It has been seen that there exists no survey paper in recent years, presenting proper literature on graph-based approaches and interference minimization algorithms, used in topology control of WSNs. Thereafter, this article has come-up with unexplored literature part related to power optimization in topology control using graph-based approaches, explicitly designed interference models and interference minimization algorithms.

5.2 Previous survey on routing techniques in WSNs

Bhattacharyya et al. in 2010 [66] offer the detail description of eight hierarchical routing protocols that are LEACH, TEEN, APTEEN, PEGASIS, SPIN, DD, RR and GEAR. Performance analysis of the mentioned protocols in term of mobility, network lifetime, power management, scalability has been provided along with the direction of improvements which can be done in future work. Review of hierarchical routing protocols from year 2011 to 2016 has been done by Sabor et al. in 2017 [68]. Sabor et al. classify the protocols into classical and optimized routing protocols dealing with mobile WSNs. Working details, advantages, drawbacks and performance analysis in respect of delay, network size and energy efficiency of the protocols dealing with the mobility of sensor nodes and/or mobile sinks have been provided. Authors have drawn the conclusion that centralized routing is more energy-efficient than the distributed routing so semi-distributed or semi-centralized protocols should be considered for future work. Also clustering techniques from year 2002 to 2016 are studied by Ogundile and Alfa in 2017 [69]. Ogundile and Alfa bring out the classification of routing

techniques engaged with mobile/static sink adopting either single hop or multi-hop routing to communicate to sink node. As per the classification, detail of the working protocols with respect to CHs selection, number of CHs, CHs rotation, cluster formation, intra-cluster and inter-cluster routing have been stated. Authors offer the future direction for more load balancing energy efficient cluster-based routing techniques. Guleria and Verma in 2019 [70] present a systematic study of hierarchical routing protocols from year 2012 to 2017. Moreover, a deep analysis of both classical and swarm intelligence (SI) based clustering techniques (PSO, ACO, ABC, BFO) opting different strategy and working methodology using performance metric of energy efficiency, data aggregation, location awareness, QoS, scalability, load balancing and fault tolerance has been provided. Enhancement of LEACH protocols are stated along with pros and cons. Authors bring out the long discussion of advantages, existing issues and improvement for future works in this field. Thereafter Fanian and Rafsanjani in 2019 [9] provide a comprehensive survey of cluster-based routing protocols from year 2000-2018. The aim is to cover all those methodologies which are energy efficient and help the readers to build their base knowledge about the methodologies opted in this domain. Authors classify the protocols on the basis of different methodologies as classical approaches, fuzzy-based approaches, meta-heuristic based approaches, hybrid meta-heuristic and fuzzy-based approaches. They put forth desired objectives, measuring parameters, merits and demerits of different routing protocols in each classification. A comparison of different routing protocols of each category has been shown on the basis of mobility, cluster size, network type, methodologies etc. Analysis of techniques on the basis of solution complexity is missing in the article. Chan et al. in 2020 [20] provide the insights on classical hierarchical routing protocols following the division into chain-based, tree-based, grid-based and area-based routing topology formation. Authors mention advantages, disadvantages and comparison of the routing techniques with respect to energy efficiency, load balancing, delay, scalability, network longevity and reliability. Moreover, study of recently developed versions of low-energy adaptive clustering hierarchy (LEACH) routing protocols are given along with comparison among the proposed versions. Authors put forth present challenges with direction of future work. Then Rawat and Chauhan in 2021 [71] present a survey of clustering based approaches by collecting efficient protocols from year 2000 to 2020. This is the most recent survey paper in the domain of hierarchical and/or clustering routing techniques in WSNs. A taxonomy has been given by classifying clustering approaches into homogeneous network-based, heterogeneous network-based, fuzzy logic based and heuristic based protocols. Authors explains methodologies of each clustering protocol with its objectives, key features, merits and demerits along with future scope. Analysis of clustering protocols have been conducted by considering the parameters of CH selection, CH rotation, data transmission, mobility, topology, routing techniques and simulator used while the comparison between different techniques are based on prototype used, energy consumption, network lifetime, throughput and stability. Ketshabetswe et al. in 2019 [6] present the working details and analysis of routing protocols concerned with energy saving and network lifetime particularly. Authors follow the classification of routing protocols based on homogenous and non-homogenous network which are further categorised based on mobility of sink nodes. Insight of WSNs model in term of graph-based structure and energy dissipation model with good explaination has been provided. Based on the classification, authors cover large number of flat-based, location-based, hierarchical, QoS aware, cross-layer, nature-inspired, cooperative and opportunistic routing protocols. Analytic comparison, existing challenges, direction of improvement and new work direction are offered in this paper. Then Lin et al. in 2020 [10] cover the energy efficient strategies for static WSN from year 2002 to 2019. They also proposed the mathematical definition of energy efficiency first time in history. A review of previously existing survey is mentioned along with drawbacks. Authors classify energy efficient strategies into five categories, the Energy-Efficient Media Access Control (EEMAC) protocol, the Mobile Node Assistance Scheme (MNAS), the Energy-Efficient Clustering Scheme (EECS), the Energy-Efficient Routing Scheme (EERS) and the Compress Sensing-Based Scheme (CSS) and put forth characteristics and applications domain of each classification. One can see open challenges, opportunities and future directions related to energy efficient WSNs. Pantazis et al. in 2013 [67] present the broad classification of routing techniques on the basis of network structure, communication model, topology based and reliable routing. The study of flat and hierarchical routings are found in network structure category, location based in topology based category and QoS based in reliable routing schemes. The aim of this review article is to assemble all routing techniques which are energy efficient as well as facilitate load balancing among the sensors. Moreover, authors present merits and demerits along with the performance of the routing techniques using the parameters of scalability, mobility, robustness, route metric and periodic message type. Cao et al. in 2019 [55] bring out the new classification of routing protocols based on transmission path, network structure, working mode, next hop routing and data delivery model. The main focus of the paper is to present literature of SI based routing techniques used in application engaging mobile WSNs. Authors systematically provide the description, discussion, analysis by stating advantages, limitations and comparison of PSO algorithms, ACO alorithms, artificial fish swarm algorithm (AFSA), artificial bee colony

(ABC) algorithms, shuffled frog leaping algorithm (SFLA) proposed in the history to overcome the complex challenges of mobile WSNs related topology designs and routing. Further Zagrouba and Kardi in 2021 [72] present a board taxonomy by classifying routing techniques into 9 different categories based on application type, delivery mode, initiator of communication, network architecture, route discovery, network structure, protocol operation, next-hop selection, latency-aware and energy-efficient routing. This paper provides the literature of real-time routing protocols under the category of application type and delivery mode, data-centric routing protocols under network architecture, location-based and hierarchical routing protocols with swarm intelligence based techniques under network structure. Performance evaluation has been shown using the parameter of reliability, congestion control, energy efficiency, data aggregation, scalability, delivery delay, network lifetime, fault-tolerance and QoS. Liu in 2017 [74] is the first who provides the comprehensive survey with new classification of ACO based routing techniques in WSNs. Author offers the working details, performance advantages and limitations of some well known ACO-based routing algorithms along with comparison and future scope. Al Aghbari et al. in 2019 [5] present the study of PSO, ACO, FA, GA, ABC, BFO techniques based routing algorithms from year 2010 to 2019 in solving WSNs problems of network lifetime, energy consumption and throughput. Authors put forth concluding remarks for each of routing techniques by stating overall performance advantages and drawbacks. According to AL Aghbari, PSO technique is optimal for cluster-based routing and converge faster than any other SI based algorithm. Also PSO consumes lesser energy compared to other mentioned techniques. Secondly, ACO based routing algorithms are more reliable for multi route data transmission and maximize network lifetime. Also Chen at al. in 2020 [75] cover the literature of ACO-based routing techniques for both static and mobile WSNs. Authors provide the methodology and performance analysis of some typical ACO routing techniques using the parameter of energy utility, load balancing and complexity. Moreover issues related to the topology of mobile WSNs are discussed with direction of improvement. Rachamalla and Kancharla in 2013 [73] offer the insight of many real-time routing protocols by providing working objectives, details along with performance advantages and drawbacks for both static and mobile WSNs. Comparison among the proposed protocols have been done using parameters of energy efficiency and reliability. Further Kim et al. in 2017 [19] discuss and analyse the real-time communication based applications in WSNs and respective routing protocols. The paper covers the literature since 2010 and offers the classification of real-time routing protocols into soft, hard and firm based on two levels, first one is MAC and scheduling and second is routing. Motivated from the previous work, this article presents the review of most popular, unexplored and latest network structure based classical hierarchical routing protocols, topology aware optimization techniques based PSO and ACO routing techniques and delay-sensitive application based real-time routing protocols. Analysis of each technique put forth the various merits and demerits related to routing issues, solution complexity, mobility/ immobility of network.

6 Review of Topology control methods

Survey papers from the previous years cover the major literature of topology control in WSNs involving the graph models, power mode based topology planning, distributed and centralized power control algorithmic approaches which we have already discussed in Sect. 5.1. However, no survey paper presents the complete study on graph based approaches in reducing and managing nodes transmit power, explicitly designed interference models and interference minimizing algorithms in TC of WSNs. This section presents the review of unexplored graph based solutions, interference minimization models and related algorithms which are used to conserve nodes power, extend network lifetime, eliminate redundant transmission and transmission cost in directed/ bi-directed connected graph models of WSNs.

6.1 Transmit power management based topology control

Topology control issue regarding energy consumption, related to minimize nodes power consumption as nodes are power constraint and depletion of their power make them die soon hence it effects network lifetime adversely. All those attempts, related to minimize and balance nodes power consumption by determining an appropriate network topology which is energy efficient and cost efficient, results an optimal power assignment to each node belonging to the topology or allow nodes to adjust their power level to design their own topology, come under 'Transmit power management based TC'. Cheng et al. in 2003 [27] formulated a 'Strong Minimum Energy Topology (SMET)' problem which is related to transmit power/ transmission range assignment to each node in the network. **SMET** states that for a given set of sensors, compute the transmit power of each sensor such that there exists atleast one bi-directional link between every pair of sensors and total transmit power gets minimized. Total transmit power is the linear sum of transmit power of each sensor node. Cheng et al. proved the NP-completeness of SMET in bidirectional graph model of WSNs and proposed two heuristics algorithms to find the approximate solution of

the problem. Santi in her survey paper [13] covers the literature related to **'Range assignment problem'**, the problem is same as SMET with the only difference is that the graph model can be directed as well as undirected/ bi-directed. In this paper, latest articles on the same problem or some extended problems from year 2012 have been reviewed and documented in the Table 1. A proper study of each technique is given in the table like the objective/ contribution of the proposed work, detail of methodology opted and limitations with its future scope. Following are the some defined terms/ definitions whose abbreviations are used in Table 1.

DPMP (Dual power management problem): Rong et al. in 2004 [76] formulated DPMP and proved the NP-hardness of the problem. It can be said that DPMP is the extended problem of SMET, dealing with optimal assignment of transmit power to each node from the discrete set of two power levels i.e. $\{H_p$ (high power), L_p (low power) $\}$. **DPMP** states that for a given set of sensor nodes and two power levels, H_p and L_p, find the optimal assignment of these two power levels to the nodes of induced graph model of WSNs such that the total power consumption gets minimized. Once the transmit power either high or low has been assigned to the nodes, it kept on fixed. Generalising the DPMP to k-distinct power levels assignment problem, named as **K–Distinct SMET** which put forth problem of optimal assignment of a transmit power to each node in the network from the discrete set of k distinct power levels such that the resultant network is strongly connected and total power consumption gets minimized. Some authors [78] come-up with fault tolerance in power management problem by introducing k-edge connectivity concept of graph theory. Going through each technique of Table 1, following are the key highlights:

- Although the developed work has provided the energy efficient topology according to the desired objective and has lesser complexity, yet is not suitable in dynamic WSNs. Topology gets disconnected with the nodes movement hence they are not adaptable to nodes mobility as once the power is assigned to the nodes, it remains fixed in the network during data transmission.
- Power assignment algorithms are limited to discrete set [24, 77, 79, 80] and the one with interval power assignment [81] is restricted to those WSNs only whose underlying graph can be an interval graph. This work needs to be generalised.
- Issues related to fault-tolerance of topology has been resolved by undertaking edge-connectivity [78]. It can be further taken to vertex-connectivity with the consideration of node failure.
- TC methods lag behind in term of network heterogeneity, scalability, fault-tolerance, mobility and routing delay.

6.2 Interference minimization approaches in topology control

Rickenbach et al. in the survey paper [14] in year 2009, present the literature of interference minimization explicitly through various interference models and algorithmic approaches. After this, no review has been conducted on these approaches. Literature from year 2009 onwards has been covered in this survey paper. Articles concerned with minimization of maximum interference, total interference and average interference have been reviewed and documented in Table 2 so that the readers can get acquainted about the contribution made by proposed work, detail of opted methodology/ approach and existing limitations. Sender-centric and receiver-centric interference models are the most opted models to minimize nodes/ links interference. Panda and Shetty in [15] come-up with new models namely SUM and MAX, to compute coverage based link interference i.e. $cov(uv)$ under SUM model is $\sum\{cov_{uv}(u) + cov_{vu}(v)\}$ and under MAX model is $max\{cov_{uv}(u), cov_{vu}(v)\}$. Consider an illustration given in Fig. 4, link interference under SUM model comes out to be 13 and under MAX model is 7. After this, Sun et al. in [37] introduced another link interference model i.e. $I_{uv} = |\{xy \mid xy$ is any link, x or y or both $\in$ Disc $(u, d(u, v))$ or Disc $(v, d(u, v))\}|$. This only computes number of interfering links instead of nodes which come inside the transmission disc of any node. This new link interference models outperform the previous models as being more practical [37]. These interference minimizing models results minimum interference topology with the issue of network sparsity and fault-tolerance. Then Bao and Deng in [38] introduced fault-tolerant and reduced interference topology by introducing k-disjoint paths between every pair of nodes of minimum interference. However, the topology formed using these techniques is keeping high latency factor. Therefore Wu et al. in [88] propose polynomial time algorithms to result delay constraint and minimum interference topology. But the assumption that all links are having same transmission power limits the use of the work. Lakshmi et al. in [17] minimize sender and receiver interference using two distinct power levels; nodes either assign with low power level or high in such a way that the formed-MST causes minimum of maximum interference. However, the proposed techniques and algorithms lag their functionality in mobile WSNs as they are designed for static network only. Furthermore, work on interference minimization can be extended to nodes with multiple power levels or continuous range of power level like an interval, to enhance the fault-tolerance and avoid energy hole problem in interference minimization, concept of examination of link quality and nodes residual

Table 1 Overview of Graph-based approaches in 'Transmit power management based topology control'

Reference	Contribution	Approach/Methodology	Limitation/Future scope
2012 [77]	Authors proposed a fully distributed power adjustment algorithm named as BRASP to form a connected and reliable topology of minimum cost by undertaking the factor of lossy links by computing link qualities locally. Worst case time complexity is $\Leftarrow\|E\|$) and cost is $\Leftarrow\|E\| + \|V\|$). Outperform other Deterministic TC algorithms in term of network energy efficiency and connectivity with decreased node degree.	Among the $K-$ distinct power levels, each node starts with minimum power level and gradually enhance its power to any of the available power level to satisfy the connectivity constraint. Connectivity constraint is the value between 0 to 1 such that the value more than 0 between the the pair of nodes depicts their connection and closer the value to 1, stronger is the connection. Authors used Dijkstra's algorithm to find reliable path of minimum cost between the pair of nodes by following the restriction of hop counts means if 2-hop restriction is imposed, nodes can adjust their power to reach 2-hop away nodes only.	Defined for static network only. BRASP will consume more energy when larger threshold of hop count is used. Work on improving end-end delivery by accounting time-dependent lossy links might be undertaken in future.
2015 [78]	Introduction of fault-tolerance in DPMP through $k-$edge connectivity, named as $k-$EDP. Authors proved the NP-hardness of $k-$EDP in geometric case and provided a 2-approximate solution to the problem in polynomial time. Moreover, proposal of 'Minimum-Cost-H Path' algorithm which obtains 2-edge connected spanning subgraph.	Author formulated the problem as LPP and proposed 2-approximation algorithm using the iterated rounding method proposed in [82]. 2-approximation algorithm keeps on applying 'Iterated rounding technique' on LPP to convert its fractional solution into integer until a k-connected graph is obtained. Minimum-Cost-H Path algorithm successively adds H-paths with the idea of picking low power edges first to obtain 2-edge connected spanning subgraph H.	Suitable for static WSNs only. The problem can be concerned with k-vertex-connectivity undertaking the node failure in future.
2015 [79]	Proposal of 2-approximate solution to DPMP by introducing SCDPA (strongly connected dual power assignment) algorithm and an algorithm concerned with the assignment of 3-power levels which are low, middle and high power level following the similar steps of the former 2-approximate algorithm.	MST of the induced graph is found by using Prim's algorithm and transmission range is assigned to each node using distance factor in MST. Now nodes whose transmission range is less than or equal to low power level is assigned with L_p and remaining with H_p.	Designed for static network only. No concern of topology fault-tolerance has been shown. A work to have fault-tolerant and minimum delay topology can be explored in future.
2015 [80]	Provide the study of NP-completeness of K-Distinct SMET problem. A hierarchical structure of K-Distinct SMET is introduced, named as 'k-Distinct-rSMEHT' problem and an approximate solution is provided which reduces total power consumption. Moreover, a polynomial time solution has been shown for the 2-Distinct-SMET problem.	Methodology of algorithm is using extension tree concept and idea of joining r-distinct spanning tree to form a single spanning tree of a complete weighted bi-directed graph model of WSNs.	Proposed solution is not optimal and suitable for static network only.

Table 1 (continued)

Reference	Contribution	Approach/Methodology	Limitation/Future scope
2016 [83]	Authors proposed topology adaptation algorithm (TAA) and preprocessing algorithm (PA) to obtain minimum power consumption and connected bidirectional routing topology in Heterogenous WSNs with Time complexity $\Leftarrow n3log^2n$). This work is extendable to different types of nodes having different power levels. More fault tolerant than conventional sparse topology.	There are super nodes and ordinary nodes present with a sink node. Each super nodes are supposed to keep k-disjoint paths with sink nodes and ordinary node keeps a single path to sink node using graph transformation and max-flow theory. TAA algorithm follows MST initially to establish the connectedness among all the nodes and sink, then more links have been added to MST to achieve $k-$connectedness of super nodes. After this graph transformation takes place to convert node-disjoint paths to edge-disjoint paths using max flow theory and best k-paths have been kept. PA algorithm is used after TAA to remove unnecessary long links without disturbing connectivity if required.	Super nodes makes the network quite expensive. Proposed work is suitable in static network only.
2017 [84]	Authors proposed two algorithms for the SMET problem, first one is the improvisation of Prim-incremental heuristic in term of time complexity and second is the hybrid of Boruvka and Prim algorithms. The hybrid heuristic is more energy efficient and faster than the previous versions of incremental heuristic on the planar graph.	Methodology of first algorithm joins the nodes in greedy fashion of increasing total power consumption to form a spanning tree. There exists a set for every node to choose its closest neighbour directly based on the increased power level. Second algorithm forms MST using the hybrid methodology of Prim and Boruvka algorithms. Each vertex is taken as a single component and edge of least power consumption is selected from each component to form a new contracted graph. Selection of edges from contracted graph is done to form MST following Prim's algorithm.	No approximation bound on optimal solution has been provided. Algorithms are restricted to planar graphs only. Also nodes are assumed to be immobile in behaviour.
2017 [85]	Proposal of an integrated power adjustment and clustering based topology control method known as 'Collaborative Topology Control Protocol (CTCP)'. This approach reduces power consumption and improves stability factor in large heterogeneous WSNs. This approach causes less overhead in topology setup phase.	The methodology starts with minimum connected domination set (MCDS) theory to select dominated nodes as CHs and establish a base topology of CHs. Thereafter a power adjustment algorithm has been initiated to assign distinct transmission range to each MN within a group by its CH as each CH keeps topological map of its surrounding. Transmission range of each node can be varied to achieve desired topology as smaller range of MN causes many CHs of long transmission range which causes more energy cost and long range of MN reduces energy cost caused by CHs. with the increment of hop distance.	This method reduced hop distance at the cost of keeping dense network which causes network interference. The proposed work is not suitable when nodes are mobile since it complicates the working of algorithms in topology establishment.
2018 [24]	Authors improved approximation bound of achieving optimal solution of DPMP and proposed a polynomial time algorithm which uses the concept of second hamiltonian cycle.	The methodology is using the contracted set theory with second hamiltonian cycle effects minimum number of nodes with high transmit power on an induced directed graph model of WSNs to obtain connected subgraph.	Designed for static and directed network. The work can be extended to mobile topology with bi-directional links.

Table 1 (continued)

Reference	Contribution	Approach/Methodology	Limitation/Future scope
2019 [86]	Authors proposed two polynomial time algorithms to obtain the solution of DPMP under two cases, 1. Power levels are chosen from the given discrete set and 2. Power levels are taken as input. Lower and upper bounds on optimal number of sensor nodes assigned with high power for DPMP have been obtained. The second algorithm performs better than first to minimize total power consumption with connectivity constraint.	First algorithm starts with the finding of MST and then assignment of transmit power to each node in MST. H_p is taken maximum of assigned transmit power to maintain the connectivity and remaining transmit powers work to find L_p. Each node whose transmission range is less than H_p would be assigned with low power level L_p, so this way minimum number of sensors is assigned with H_p and total power gets minimized in the connected network. The methodology of second algorithm is to assign low power to all the nodes inducing a spanning forest which consists of one or more components. Assignment of high power is done to only those nodes which effect maximum number of connected components.	The obtained solution is 2-approximate, lacking optimality. Both the algorithms are suitable for static network only.
2019 [87]	Proposal of an virtual topology based routing protocol where nodes can vary their transmission range. The proposed algorithm is distributed in nature and improves network life, routing cost and having polynomial time complexity. This is better than the previously existing CDS based techniques.	In this, authors modelled the WSN as a disk graph and engaged minimum connected dominating set (MCDS) to have the connected topology among the nodes which results quick search of routes with shorter distance by keeping minimum disc radius and variable transmission range.	This approach is not much suitable for directed and mobile WSNs. Even though nodes can have variable transmission range but the effect of topology construction on nodes power level didn't show explicitly.
2019 [17]	Formulate a k-power level assignment problem and provide a near-optimal solution by proposing two polynomial time algorithms. First is MST-based and second is Incremental heuristic. Incremental heuristic performs better in minimizing total power consumption than MST-based.	The methodology of first algorithm starts with the set of k-power levels in ascending order and computation of MST. According to the weight on each edge on MST, transmit power from the set of k-power levels are assigned using divide and conquer rule i.e. if weight of node u is lying between p_i and p_{i+1}, then u is assigned with p_{i+1}. The methodology of second algorithm forms a spanning tree with one by one addition of edges in increasing order of their weights. The difference from the former algorithm is that this algorithm maintains two sets for vertices, one shows visited vertices and other unvisited.	In MST-based algorithm if weight on any edge is more than the available power level then no feasible solution exists and algorithm gets terminated. Also it is designed for static network only.
2021 [81]	Proposal of an interval graph based routing topology which minimize nodes' energy consumption and results an appropriate assignment of power level to each node while ensuring topology connectivity. This is the first approach towards WSNs to model it as an interval graph where nodes are explicitly assigned with power intervals and proposed algorithms are energy optimal and having polynomial time complexity.	Each node of WSN has been assigned with power interval ranging from minimum power level to maximum power level, required to approach different destination node. According to nodes power interval, sorting has been done to form right end interval graph model of WSN. Three algorithms have been proposed to construct an energy optimal routing topology which minimizes maximum nodes' power consumption, maximum transmission cost of worst case and total transmission cost of worst case. Methodology of proposed algorithms are based on to join maximally assigned power interval to minimally assigned power interval because transmission energy along that path is minimum.	Proposed work is suitable for static WSNs. Algorithms are centralized in nature and restricted to the sorted model of WSNs and only those WSNs whose underlying graph can be an interval graph.

Table 2 Overview of 'Interference minimization approaches' in topology control

References	Contribution	Approach/methodology	Limitation/future scope
2009 [36]	Proposal of two polynomial time algorithms MI-S and LS-IM to minimize maximum interference. Both the algorithms show good performance in some instances. Also analysis of the proposed algorithms in best and worst case scenario has been shown.	MI-S algorithm works to join edges in increasing order of their weight to form a spanning tree which minimizes maximum interference of the network. LS-IM involves local search technique to improve the solution obtained by MI-S if there is any change in network topology because of addition or deletion of edge.	Algorithms don't minimize interference optimally in every instances and is limited to 2-D topology.
2011 [89]	Proposal of first exact algorithm to minimize average interference and provide proof to the bound on maximum interference while minimizing average interference in 2-D network. Authors presented study of minimization of average interference in connected topology using no-cross property and proposed a fast exact algorithm to minimize the average interference for 1-D model in polynomial time.	Using the concept of mutual exclusion of long edges, computational geometry based on distance and transmission range, algorithm 1 obtains low interference spanning tree. Algorithm 2 considered mutual interference and right side topology for the nodes lying from left to right to construct a low interference topology.	Suitable for static nodes deployment and low dimension topology. Assumption of interference range is same as transmission range is limiting this work. Work can be extended to network where transmission range is taken more than interference range.
2011 [90]	Proposal of exact algorithms to minimize receiver-centric interference for Highway model. Multiple planar connected topologies with reduced average and maximum interference can be constructed using these algorithms.	Utilizing no-cross property to obtain connected planar topology and dynamic programming, algorithm 1 yields optimal spanning tree for a highway model.	Algorithms are designed for static network using 1-D Interference models only. Minimizing the maximum interference for the highway model is still an open problem.
2011 [91]	SBYaoGG (Smart Boundary Yao Gabriel Graph) for low interference connected network has been proposed. This optimizes transmission cost of symmetric links and resultant topology has energy efficient and low-obstacle routing.	SSYaoGG algorithm is the hybrid of GG and YG algorithms adopting smart region boundaries. This algorithm involves neighbour discovery utilizing maximum transmission range, computation of GG from the UDG yields planar topology with linear number of edges then computation of average direction vector and YG on reduced topology yields power spanner connected planar topology.	Reduced topology is sparse and every node sets to minimum transmission range reduces link reliability and quality of service (QoS).
2012 [92]	Proposal of an interference minimization technique by reducing the communication radius of nodes which therefore reduces network interference while maintaining connectivity constraint. This defined technique is applicable to asymmetric case also.	Technique works in localized way i.e. let r_{min} is the boundation set on the communication radius of each node which is actually the longest edge distance and the sensor field is partitioned into cubes of side length r_{min} which therefore results communication radius of each node inside any cube less than r_{min}.	Assumption of uniform range assignment to each node limits the applicability of the proposed technique. Also working analysis using mobile nodes has not been given.
2013 [15]	Proposal of 'Minimum Interference Strong Bidirectional Topology (MISBT)' algorithm with two new interference models, SUM and MAX, to estimate coverage-based and transmission-based nodes interference. Polynomial time MISBT algorithm under MAX model optimally minimizes maximum nodes interference and yields $2-$ approximate solution to min-average nodes interference.	MISBT algorithm involves the computation of complete weighted graph of the given WSN where the weight on each edge is found by adopting either the MAX or SUM model, then computation of MST yields minimum interference connected topology and reduced power assignment to each nodes.	Algorithms result magnitude of available Interference in integer number although adding the situations like signals fading and collision, interference should be taken as a real number, varying in nature.Designed for static WSNs. The work can be extended by allowing interference to result in a real number considering real-life applications.

Table 2 (continued)

References	Contribution	Approach/methodology	Limitation/future scope
2015 [37]	Presented a new link interference model which counts the actual number of interfering links with the given link by discarded the previously designed link interference model based on coverage. Author proposed a centralized greedy algorithm which construct an interference-optimal topology based on the newly proposed link interference model. The proposed link interference model is independent and can be applied to other topology control algorithms as well.	Working of the algorithm is to sort links in increasing order of their weights which is depending on link interference and distance, then insertion of links one by one to obtain an energy spanner connected topology. This reduces the chances of loop formation and redundant addition of links as one by one insertion of links is adopted.	Algorithm is defined for static WSN only. Individual focus on interference reduction makes network sparse brings reliability and energy hole problem.
2016 [38]	Proposed a fault-tolerance and interference-aware topology control (FICTC) algorithm which is having polynomial time complexity. FICTC minimizes maximum link interference in bidirectional network and meeting the fault tolerant requirement.	The algorithm works to achieve different k-connectivity between different pair of nodes by establishing k-disjoint paths between any pair of nodes. This is done according to obtain minimum vertex separation set with the use of nodes splitting technique. This splitting technique employs two nodes like v' and v'' in place of a node v in the original topology to fulfil the connectivity requirement. Then Prim's algorithm has been used to find augmented paths between the nodes on the resultant topology, obtained after splitting operation and updation of links capacity on the obtained augmented path. All the augmented paths merged together to obtain final k-connective topology with reduced link interference.	Defined for static network only. Adoption of k-disjoint paths between every pair of nodes causes more data transmission and so more energy consumption.
2018 [88]	Proposed one centralized and two distributed topology control algorithms which jointly work on reducing link interference and transmission delay in the connected network. Authors undertook three optimization problems: minimize maximum interference, minimize average interference and minimize path interference.	The first algorithm MDTC is centralized in nature adopted greedy strategy like LISE in which links are sorted in ascending order of their weights. Weights are the link interference and links join themselves if they satisfy delay bound in the increasing order of their weights. Second algorithm is DMST which builds a delay-constraint MST to minimize average links interference with restricted hop counts to satisfy the delay bound. DMST adds links of minimum interference whose hop counts must be less than imposed delay. A local version of DMST is also proposed by author in which nodes make next move on the basis of neighbour information i.e node choose another node if the corresponding path has minimum interference satisfying delay constraint. A polynomial time LDBF algorithm is proposed to minimize path interference. LDBF utilizes the concept of Bellman-ford algorithm in a localized way with the inclusion of transmission delay restriction over the path.	MDTC has higher time complexity. All links are assumed to have same transmission power. Delay is measured on the basis of hop counts only. Although delay is based on link quality and distance.

Table 2 (continued)

References	Contribution	Approach/methodology	Limitation/future scope
2019 [93]	Adopted the strategy of additional nodes deployment in the sensor field at those position which helps in minimization of sender interference without disturbing the established nodes position. Authors proposed a polynomial time algorithm which minimizes the maximum sender interference with preserving connectivity.	Working of the algorithm named as MSD, is to connect nodes to their nearest nodes which results each edge is Gabriel and induced a spanning forest. A MST is obtained from the resultant GG such that some edges belong to MST not to the spanning forest. Then a new node is positioned at the bisection point of that edge which further reduces the range assignment of the end nodes ensuring connectivity requirement.	Designed for static network. Deployment of more nodes causes more hops to cross which may lead transmission delay.
2019 [17]	Proposal of MST-based algorithms to minimize maximum sender and receiver interference with DPMP. Algorithms yield minimum interference connected topology and having polynomial time complexity.	Working of proposed algorithms is to partition edges into two sets i.e. high power set and low power set according to their weights where weight on each edge is defined on the basis of Euclidean distance. Edges with low power are added first to construct topology following the computation of sender (or receiver) interference of each node. Then high power vertices are sorted in ascending order of their sender (or receiver) interference and added accordingly to the topology with high power assignment until a loop free connected topology has been obtained.	This work is limited to interference minimization with dual power assignment to stationary nodes.

energy can be taken into consideration so that the existing topology could be reorganized and managed. Following are some defined terms which are used in Table 2.

- **Highway model:** It is obtained when nodes are distributed linearly in one-dimension only. It is established to obtain a lower bound on network interference.
- **Unit disk graph (UDG):** UDG models WSNs in which two nodes communicate if and only if the distance between them is at most $r \leq 1$, where r is the transmission radius which is equal for all nodes.
- **Gabriel graph (GG):** GG models WSNs in which two nodes u and v communicate iff the disk obtained using the uv as a diameter doesn't contain any point.
- **Yao graph (YG):** YG models WSNs with an integer parameter $k \geq 6$ which defines k cones. In each cone, node v chooses closest neighbour node u with the distance atmost 1.

7 Review of routing techniques

A thorough review has been conducted according to the classification made in sect. 3. Review of network structure based classical hierarchical routing protocols, topology aware optimization technique based PSO and ACO routing techniques and delay-sensitive application based real-time routing protocols, is presented using tabular format to offer the reading convenience. Each routing category is the most evolutionary, well established and quite applicable. Still there are lots of persisting issues with each categorised technique which we have discussed as well.

7.1 Classical hierarchical routing protocols

This section covers the review part of classical hierarchical routing protocols with description of basic methodology, terms and abbreviation which have been used in documenting the reviewed data in Tables 3 and 4. Table 3 presents brief overview of most popular energy efficient classical hierarchical routing protocols from year 2001 to 2016 which have been extracted from previous year survey papers [2, 4, 6, 20, 67], to offer readers idea of their methodology, merits and demerits. Table 4 presents the insights on latest classical hierarchical routing protocols from year 2017 to 2021 by mentioning the contribution made by the proposed protocol which includes the beneficial part, different approaches/ methodology opted and persisting limitations which might be overcome in future work. Research on hierarchical routing protocols have overcome the many limitations of flat based and location based routing protocols in

term of scalability, energy consumption, latency, dynamic topology, algorithmic complexity, load balancing, network heterogeneity and life cycle. The basic methodology and aim involved in this type of routing is the formation of optimal number of clusters, cluster size, appropriate selection of cluster head (CH) node in each cluster, optimal assignment of member nodes (MNs) to each CH, known as cluster member, optimal intra and inter-cluster routing. A basic network structure of cluster based/ hierarchical routing is shown in Fig. 7. In hierarchical routing, cluster of nodes can take any shape like chain, tree or grid since partition of target area has been done in many shapes like square, circles, grids and concentric circles (refer Table 4).

Some authors introduced the concept of rendezvous points (RPs)/ sojourn points (SPs) selection in cluster-based routing where mobile sinks (MSs) are engaged to aggregate data from member nodes and/or CH nodes by halting at each of RP/SP [94, 95]. Basically RPs and SPs are stoppage point for mobile sink (MS) node and authors have been dealing to find optimal RPs/ SPs to improve load balancing and reduce energy hole problem specifically. Both RPs/SPs and CHs are chosen on the basis of distance factor, member node density and energy factor. Irrespective of this, intra-cluster and inter-cluster routing are need to optimize. Some authors [96] involve time division multiple access (TDMA) mechanism to intra cluster routing where member nodes transmit data in a specified time slot to CHs and keep themselves inactive otherwise. Use of code division multiple access (CDMA) [97] to inter-cluster routing where a CH sends its data only by acknowledging code associated to it. Both TDMA and CDMA avoid unnecessary data collision, transmission, congestion at CH node or BS and consequently save nodes energy. Following are the key points, we draw from Tables 3 and 4.

- These techniques form routing topology for both homogenous and heterogenous WSNs.

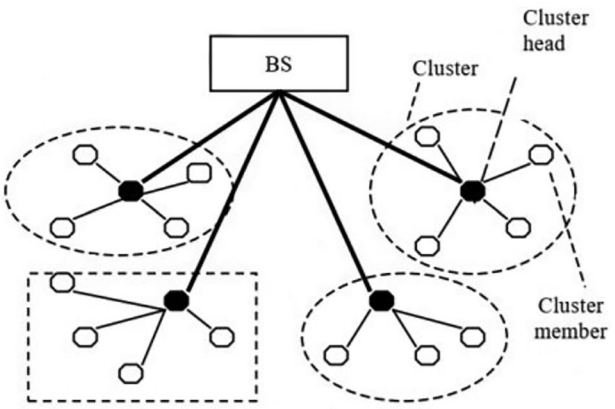


Fig. 7 Network structure of cluster based/ hierarchical routing

- Some routing protocols are suitable for both static and mobile WSNs by offering mechanism of re-clustering and re-routing and some limit themselves to static network only.
- Cluster formation makes the network more scalable and energy efficient yet they are showing limitation when timely data delivery is required. For this some author brings the concept of engaging mobile sinks (MSs) so that they themselves come to collect data from the shortest route possible but not on the urgency basis. Therefore timely data delivery is the persisting issue.
- These routing causes overhead and energy-hole problem when the hierarchy formation is chain and tree like.
- The most evolutionary part of these techniques is to have fault-recovery mechanism using beacon message exchange makes the network more reliable and fault-tolerant. However more techniques to deal with message overhead issue, fault-tolerance and energy-hole problem in dense, heterogenous and dynamic WSNs need to develop.

Moving to more stable, scalable, robust and multi-objective optimizing PSO and ACO based routing techniques.

7.2 PSO and ACO based routing techniques

From the previous section, we see that topology formed using classical hierarchical routing protocols is not robust to nodes movement and failure. Also the proposed techniques lag their performance while solving more complex multi-objective problems. So to overcome these issues and to find optimal solution of multi-objective problems, PSO and ACO based routing protocols/ techniques have been proposed for many past years. Both PSO and ACO techniques have lower complexity, more stability and robustness than classical hierarchical routing techniques. Both PSO and ACO work globally by acknowledging the whole network and setup routing topology that's why we call PSO and ACO, topology aware optimization techniques. Moving forward to look latest routing techniques using PSO and ACO. Many survey papers exist covering these techniques till the current year, so this survey paper provides the review of well established articles from year 2017 onwards and analysis of each technique is presented using tabular format by mentioning the objective/ contribution of the proposed work along with its benefits, detail of approach/ methodology opted with persisting limitations. Table 5 presents the review of PSO based routing techniques and Table 6 presents the review of ACO based routing techniques. For better understanding of PSO and ACO technique based routing schemes, an introduction of PSO and ACO technique is provided here.

Table 3 Summarization of classical hierarchical routing protocols, reviewed in previous survey papers

Routing protocol	Brief overview
Low-energy adaptive clustering hierarchy (LEACH) [98]	The first hierarchical routing protocol which is localized and distributed in nature, where nodes opt direct communication to CH and sink nodes.
	LEACH saves transmission energy and network lifetime upto small extent as CHs consume lot of energy if sink node is located far from CHs.
	Not suitable for large dynamic network as it causes data congestion at CHs and sink nodes.
	Only applicable to homogenous network since each node is assigned with same power level.
	Many modified version of LEACH have been proposed to overcome the issues concerned with energy, CHs selection and network lifetime.
Power-efficient gathering in sensor information systems (PEGASIS) [99]	Chain topology formation by sensors and sink nodes to aggregate and transfer data rather than multiple clusters formation.
	PEGASIS outperforms LEACH since it adapts to different network sizes and topologies, reduces extra overhead of LEACH by using multi-hop routing.
	Causes excessive delay due to chain topology for distant nodes and wastage of energy due to redundant transmission.
	To overcome the issues of delay and energy consumption, many modified versions of PEGASIS have been proposed as well.
Threshold sensitive energy efficient sensor network (TEEN) [100]	Hybrid of hierarchical and data-centric mechanism, reduces number of transmission and so energy consumption.
	Suitable for time-sensitive data aggregation and transmission so it is not applicable where continuous data flow are needed.
	Since the methodology of TEEN is threshold-based, the protocol causes routing complexity.
Adaptive threshold sensitive energy efficient sensor network (APTEEN) [101]	Extension of TEEN protocol, offers the modification in periodicity and threshold value of TEEN as per the requirement.
	Causes additional complexity on monitoring and altering threshold values.
	APTEEN consumes lesser energy than TEEN.
Virtual Line-based data dissemination (VLDD) [102]	A virtual line structure (VLS) has been constructed by the nodes so that lines can be acted as rails to define relay path between source to the location of MS.
	Engagement of MSs and VLS structure helps in reduce delivery time, redundant transmission and so energy consumption.
	Although frequent VLS formation causes extra overhead and long line structure may cause transmission delay in large-scale WSNs.
Virtual Grid Architecture (VGA) routing [103]	Fixed cluster-based routing topology formation where each sensor node keep static at their initial position.
	Achieves energy efficiency and maximization of network lifetime.
	Selected CHs are not optimal.
Base station controlled dynamic clustering protocol (BCDCP) [104]	In this protocol, BS plays the role of CHs selection based on energy level and each cluster shares the same number of member nodes.
	Avoids congestion at CH nodes and improves network lifetime.
Energy-balanced chain-cluster routing protocol (EBCRP) [105]	Clusters are formed by partitioning the sensing area into rectangular zone and CHs are chosen based on residual energy level. Chain connecting the CHs brings the routing path to sink node directly.
	CHs rotation results load balancing and enhances network lifetime.
	Adoption of single-hop data transmission, more energy is consumed if nodes are distance apart and so not suitable for large scale networks.

Table 3 (continued)

Routing protocol	Brief overview
Chain-based hierarchical routing protocol (CHIRON) [106]	Multiple clusters are formed by partitioning the sensing area into fan-shaped area. Nodes with higher residual energy become CHs and chain of data aggregation and routing starts from the most distant node from sink node.
	Adopting multi-hop intra and inter cluster routing, so this is suitable for large scale networks.
	CHIRON is more energy efficient, scalable and reduces delay factor in data transmission compared to PEGASIS and ECBRP.
Energy-aware data aggregation tree (EADAT) [107]	Centralized protocol since tree topology is formed by sink node where nodes with high residual energy become parent node and as the energy of parent node depletes child will switch to different parent.
	Overall path establishment is done on the basis of residual energy so it results increment in total energy consumption and longer time delays since distance factor has been avoided.
Balanced aggregation tree routing (BATR) [108]	A centralized protocol which constructs MST based on energy consumption and doesn't consider residual energy factor so it may lead transmission failure.
	Child nodes are uniformly distributed under each parent, leads to load balancing and increment of network lifetime.
Enhanced tree routing (ETR) [109]	A distributed protocol form tree-like hierarchy where each node keeps routing table containing next-hop details and path of lowest hop count is preferred.
	This protocol saves transmission energy until nodes have sufficient residual energy otherwise transmission failure happens.
A novel energy efficient clustering protocol (NEECP) [110]	In this protocol, nodes can adjust their sensing range to form clusters and select CHs.
	Adjustable sensing range is based on distance parameter between the nodes and sink node.
	This protocol is the solution of energy hole problem and minimizes the energy consumption.
Grid-based multipath with congestion avoidance routing (GMCAR) [111]	Grid-based cluster formation such that one CH is appoint to each grid for data aggregation and sending data to adjacent grid CH.
	GMCAR is good for dense and large scale routing since this offers backup CHs and multi-path data transfer if required.
	It is energy efficient, reliable and reduces transmission delay.
sleep scheduled and tree-based clustering (SSTBC) [112]	This protocol makes redundant nodes to sleeping state with the formation of tree based routing structure.
	More energy efficient than ETR, BATR and EADAT.

PSO technique PSO is an abbreviation of 'particle swarm optimization' imitating the behaviour of birds flocking in search of their food. PSO technique can solve complex multi-objective optimization problems with the feature of available swarm particles and move these particles in the search-space according to simple mathematical formula (or any multi-objective fitness function f) over the particle's position and velocity. When it comes to routing in WSNs, particle size is number of sensor nodes, say N_p and let D is the dimension of each particle, depending upon the dimension of objective function (fitness function) defined. Each particle has its own position and velocity, let X_{ij} and V_{ij} is the position vector and velocity of the i^{th} particle P_i and $1 \leq j \leq D$. Each P_i is represented as $P_i = [X_{i1}, X_{i2}, ..., X_{iD}]$.

A multi-objective fitness function has been formulated by authors to evaluate each particle quality towards the solution. Each particle iteratively improve its solution quality to converge to an optimal solution which is globally accepted. To achieve optimal solution, each particle follows its own best, called P_{best} and the global best called G_{best} to update its own velocity and position. In each iteration, velocity V_{ij} and position X_{ij} in the dimension D is updated using the following equations, respectively

$$\begin{aligned} V_{ij} = w \times V_{ij}(t-1) + c_1 \times q_1 \times (X_{P_{best}ij} - X_{ij}(t-1)) \\ + c_2 \times q_2 \times (X_{G_{best}j} - X_{ij}(t-1)) \end{aligned} \tag{2}$$

Table 4 Overview of Latest classical hierarchical routing protocols

Reference	Contribution	Approach/methodology	Limitation/future scope
2017 [113]	A new distributed unequal cluster-based routing (DUCR) algorithm aiming to mitigate coverage/energy hole problem in heterogenous immobile WSN has been proposed. DUCR algorithm utilizes energy of sensor nodes in an efficient manner, avoids the energy hole problem, improves network lifetime and tolerant capacity. It achieves constant time and message overhead.	DUCR algorithm dynamically reduces cluster's size as it requires to form more clusters near BS for load balancing and relaying. CHs adapt their cluster sizes according to their residual energies i.e. cluster's size reduces with depletion of energy over time. Nodes with higher energy selected as CHs and MNs choose their appropriate CH with the help of elementary, normalized and cost matrix. Factors involved in formation of elementary matrix are CHs residual energy, CHs distance from BS and distance of CHs from MNs. This elementary matrix is normalized for computation of cost matrix where the maximum value in the cost matrix gives the best CH to the MNs. Intra-cluster routing is done using TDMA mechanism and inter-cluster routing requires CHs to distribute their data using a cost function to the lower level CHs on receiving beacon messages. If they are not available CHs directly do the transmission to BS maintaining network connectivity. DUCR involved fault-recovery mechanism by removing dead node, re-clustering and re-routing with the help of nodes local information.	No control on number of CHs. Multi-hop inter-cluster routing causes latency. Designed for static and small scaled network only.
2017 [114]	Authors proposed a cluster-based approach to control the transmission power problem and energy hole in strip-based WSNs. The proposed work is the very first contribution to the strip-based model of WSNs which has applications in mines, railway lines, bridges etc. The method improves network lifetime and offers the optimal relationship between transmission distance and node position.	The methodology engages accurate distance based division of target region into different layers as clusters and each layer keeps a CH. CH keeps on change as per rotation of data transmission within a layer. Intra-cluster routing has been done in single-hop manner and inter cluster routing can take multi-hops. Total energy consumption of every layer has been computed and in order to maximize life of CH of the layer, closer to sink, author brings optimal length of each layers and transmission distance as layer closer to sink node should has smaller length so that the corresponding CH carries smaller transmission range to extend its life.	Assumption of static sensor nodes and initial power levels of each node must be same except sink node limits the application of this work.
2017 [115]	proposed an 'energy aware path construction (EAPC)' algorithm to further improve path efficiency for data collection, prolong network lifetime and reduce energy consumption.	The methodology starts with the formation of MST in the target area and thereafter MST has been partitioned into disjoint tree like structures with the root node is CP. There are as many trees as many CPs and each CP collects information from rest of the nodes belonging to its corresponding tree. MS builds the trajectory to visit each CP in such way that the most burdened CP must be visited first while satisfying distance constraint, buffer constraint and battery constraint.	Assumption of static sensor nodes and mobile sink having same communication range limits the applicability of the proposed work.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2018 [116]	Authors proposed a duty-cycle based clustering routing protocol. This approach minimizes transmission energy, improves network lifetime, reduces unnecessary data transmission and collision of data.	This approach is consisting of four main phases: clustering, sleep/awake scheduling of nodes, two level data aggregation and inter-cluster and intra-cluster routing. Initially weight of each node is computed based on nodes energy level and node degree and then clusters are formed using Voronai diagram. Nodes with highest weight becomes CHs and based distance each CH chooses its group member. In each cluster, nodes except CH are set to sleep and awake cycle using MAC protocol by assigning appropriate time slots. Data aggregation is done by intermediate node as well as CHs based on the corresponding node position. Intra-cluster routing used a 'hybrid chicken swarm optimization algorithm' which reduces redundant data transmission and compress the data to transmit via optimal route. Inter-cluster routing has been done by constructing tree structure based on sink node position.	Defined for static and small network only. Authors didn't provide the complexity of the proposed work. This work can be extended to dense network in future.
2018 [117]	A clustering-based routing protocol for dynamic WSNs (CRPD) has been proposed. CRPD supports updation of network topology in case of dynamic addition of new nodes and their partial mobility in energy efficient way. This protocol has constant message and time complexity.	Network topology gets updated periodically by doing re-clustering and re-routing with the nodes mobility and energy depletion of CHs. Only nodes which is having more than 40% of their initial energy can compete for CH position and among those, the node which is having maximum degree chosen as CH. Also there can be more than one node receiving *youarecheadmessage* in the same ngh, for this, author introduced the intelligent waiting policy in which nodes have to wait for certain time before taken itself as a CH as they may receive *iamcheadmessage* from the actual CH within a given time and this makes selection sorted. Author used *ack* and *ACK* confirmation messages for reliable data transmission.	Immobility of sink causes data transmission delay. This method is defined for homogenous WSN only. The network lifetime of CRPD decreases with the increase of network size.
2018 [118]	Proposal of a 'Cluster-head Restricted Energy Efficient Protocol (CREEP)' concerning heterogeneous WSN. The focus of CREEP is to enhance network lifetime and throughput. CREEP lowers system complexity by limiting no. of CHs and performs better as compared to other existing protocols in heterogeneous environment for both stationary and mobile WSNs.	In CREEP some nodes are assume to be advanced and rest are normal. A distance component is introduced with residual energy factor in selecting CHs to ensures that the nodes far-away from BS get lesser chance of becoming a CH. CREEP approximates the square WSN field to a circular field as a disc of radius R and BS at its centre. CHs inside that disc adopt single hop routing while others use dual-hops. CREEP puts the limit on number of CHs nodes by introducing threshold on number of alive nodes inside the clusters.	CREEP has been designed for static network with BS located at centre, limits its applicability. Algorithm can be improved by including spatial distribution of nodes during CH selection process.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2018 [119]	Proposed a centralized 'hybrid energy efficient multi-path routing protocol (HEEMP)' for heterogeneous WSN. This offers multiple path availability which makes network more reliable. Also HEEMP increases network lifetime.	In HEEMP, sink does all network management task by keeping every node information, select CH nodes by calculating 'Chance of election (CE)' parameter based on normalized value of residual energy and node degree. Sink initiates multi route construction after CHs selection in which it forms a 'Route Set', containing all possible routes from CHs to itself and then reduces it to 'Legitimate Route Set', consisting of those routes in which nodes have sufficient level of residual energy. Intra-cluster routing is done either in 'Philanthropist mode' or 'selfish mode'. In philanthropist mode, node transmits to one of its neighbour node which is having maximum remaining energy whereas in selfish mode, node only aims at reducing its own energy expenditure by sending packets to the nearest progressive node.	HEEMP relies on sink to compute the topology, which may result in longer delays. Assumption of immobile network and error free communication restricts HEEMP applicability.
2018 [97]	Proposal of an unequal clustering approach to mitigate energy hole problem using multi-hop routing, named UCR-H. This approach is suitable with heterogenous energy level of nodes and minimize energy consumption with the improvement of network lifetime.	The target area has been partitioned into equal size rectangular area and distance between the nodes has been estimated with the help of received signal strength. Number of clusters and cluster size has been chosen based on distance from BS and network density. There are more clusters of small size nearer to BS to balance the burden among the CHs. The node with highest power level is chosen as CH in heterogenous WSNs and rotation of CHs happens after reaching below a power threshold. Intra cluster routing follows TDMA schedule and Inter cluster routing follows CDMA code.	BS are placed outside the rectangular target field and all the sensor nodes are assumed to be fixed, limits the applicability of UCR-H.
2018 [120]	Authors proposed a virtual-grid infrastructure and mobile sink based routing algorithm. This approach reduces hot-spot problem, end-end delay and balances energy consumption as compared to other similar algorithm. The proposed work is suitable for small as well as large WSNs. Also this doesn't need trajectory definition of MS and can operate for multiple MS.	In this, network has been partitioned into square shaped cells of equal area and nodes closer to partitioned line, called intersection nodes, save the position of mobile sink node after every updation. Rest of the nodes keep the location information of their neighbour nodes including intersection nodes belonging to the same cell and based on the distance, nodes transmit their data to intersection node using minimum hops. Intersection nodes transmit data to sink node if they satisfy threshold based on Euclidean distance and residual energy. In case of failure of intersection node, a neighbour node of minimum distance, satisfying threshold is selected as an alternate intersection node using beacon messages.	Designed for static and uniformly distributed sensor nodes. Since the proposed algorithm is defined for 2-D model based. This can be extended for higher dimension.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2018 [94]	Proposal of a distributed data gathering technique consisting of cluster-based RPs selection algorithm and three trajectory design algorithms for MS with three different objectives: reduced energy path (REP), reduced delay path (RDP) and delay bound path (DBP). REP generates an energy-efficient trajectory that avoids the sensor nodes from using multi-hop communication. The RDP follows the same technique with delay reduction by cutting of REP traced path. The DBP algorithm determines a time restricted trajectory. These algorithms increase network lifetime.	COST and WEIGHT are defined parameters for RPs selection, the sensor nodes having highest COST among their ngh nodes, announce themselves as RPs and rest of the nodes join RPs in their ngh of lowest WEIGHT. After this, path formation for MS is initiated where MS has to visit each RPs for data collection. Author has given three types of path formation algorithms, 1. REP which determines path of minimum energy consumption. Author used Christofis's heuristic here. 2. RDP which determines delay reduced and energy efficient path. In this, some of RPs are excluded from the path if their transmission range disc intersecting with any edge of the acquired path. 3. DBP designs a delay sensitive path by restricting path length travelled by MS. If the designed path's total length exceeding the defined threshold then one or some RPs need to eliminate which falls in the middle of two adjacent edges of maximum path length sum.	RDP consume more energy in data collection and transmission to the MS. Assumption of MS has enough energy to gather data from all the RPs restricts the applicability of protocol.
2018 [121]	Authors proposed a 'Grid-Cycle Routing Protocol (GCRP)' using mobile sink to minimize hot-spot and overhead problem caused by location updation of MS. This approach improves network lifetime and timely data delivery. This approach is more scalable and stable than the existing grid-based routing protocols.	GCRP partitions the target area into square shaped grid cells, each cell is like a cluster and in each grid, grid cell head (GCH) like CH has been selected based on distance from the cell centre and residual energy. Joining of the all GCHs form exterior cycle where the boundary of each GCH in involved and interior cycle where no GCH boundary is involved. For data aggregation, MS rotates around the GCHs and collect data from the closest GCH. Reselection of GCH has been carried out when the existing GCH energy level gets lower than the defined threshold. Only a closet GCHs update themselves with MS position and circulate the information.	Designed for static and homogenous network. Also assumption of obstacle free path between two nodes makes the proposed work impractical.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2019 [96]	Proposal of 'cluster-based fault-tolerant routing (CFTR)' protocol to mitigate frequent re-clustering which reduces overhead as well as extra energy consumption to tolerate nodes failures. CFTR is suitable for heterogeneous large-scale network, having constant time and message complexity.	Working of the protocol is to have higher energy nodes as CHs and operate multiple rounds without repetitive clustering. A cost function is defined having parameters: Energy Cost (EC) and no. of cluster members in the previous round (NCPR) for MN nodes to select appropriate CHs. There exists a case where some nodes are not able to find any CH within their communication range, CFTR sorted this situation by engaging multi-hop scheme, appointing some nodes as relay nodes and establish connection to an approachable CH. Intra-cluster routing is done using TDMA schedule and inter-cluster routing requires CHs to route their data via other CHs using a cost function that considers routes with sufficient energy for data forwarding and load is evenly distributed among the CHs. Further author introduced a low overhead fault-tolerant mechanism to cope up with CHs failures by employing remaining CHs nodes and normal nodes as backup nodes.	Designed for static WSN only. Limitation on optimal number of CH nodes and recovery strategy can be done only on CH nodes.
2019 [122]	Proposal of an energy saving routing approach by integrating geographic routing with hierarchical and using mobile sinks, abbreviated as IoGHR. This approach increases network lifetime, decreases energy consumption as well as latency factor.	Based on the nodes density, target area has been partitioned into equal size grids. Each grid consists of 4 MNs and 1 CH. CH has been selected based on highest power level or residual energy. Sink node sends its location to each CHs and moves using time schedule, this minimizes message overhead and data collision since each CH is aware of sink time schedule and so pass their data at appropriate time. Intra cluster routing can use maximum 2-hops for data transmission.	The proposed work is not suitable when nodes are mobile. The work is extensible to multiple MSs but this makes the network quite costly.
2019 [123]	proposal of a cluster-based 'energy-efficient scalable routing algorithm (EESRA)'. The proposed work extends the network lifetime in a large scale network which is a drawback of LEACH protocol.	In EESRA, each CH keeps cluster congregation (CG) nodes to aggregate sensed data from the CMs using CSMA manner and pass their data to the corresponding CH using TDMA manner. Thereafter CHs transmit data to BS. That's how EESRA forms three-layer hierarchy topology to reduce the CH load.	Assumption of all sensor nodes are homogeneous and fixed after deployment in a square two-dimensional space limits the applicability of EESRA. Moreover, the space is assumed to be free of obstacles is quite impractical.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2019 [124]	Authors proposed an 'energy-balanced tree-based data collection (ETDC)' strategy using MS. ETDC reduces transmission cost with end-end delay and prolongs network lifetime. ETDC is suitable in large-scale network as well. ETDC has polynomial time complexity.	Target area has been constructed as a tree with centre node taken as a root of data gathering tree. Nodes which are closest to the root node have the highest priority to select as child nodes. This process of child and parent formation keeps continue until all the nodes get covered to form a tree. In the process of building the data collection tree, nodes need to frequently communicate with each other by estimating signal strength and the distance between them. The constructed tree has been partitioned into distinct subtrees like clusters and MS gathers data from each of subtrees by optimizing moving trajectory of MS by using least square method.	Designed for static WSNs. This is centralized approach, operated when center node sets itself as a root of data gathering tree. Also there can be redundant data gathering possible within each subtree in case of big data.
2020 [125]	Proposed an 'Iterative Sensor Clustering and Mobile Sink Trajectory Optimization (ISCTO)' technique aiming to reduce energy-hole problem in non-uniform nodes distributed WSN with low computational complexity.	An optimization problem considering total energy consumption as objective function is formulated. To reduce computational complexity, considered problem is decomposed into two optimization problems: one is sensor clustering problem, considering the energy consumption of the SNs and other is trajectory optimization, considering the energy consumption of multiple MSs. ISCTO considers density function and residual energy for CH selection. SNs which can minimize defined density function and having higher residual energy are chosen as CHs. Authors introduce a sector boundary control method in which network is divided into several sectors to minimize the amount of energy consumed by multiple MSs in travelling and data collecting.	Specifically designed for network where sinks are mobile and other nodes are immobile. Engaging multiple MSs to eliminate hot-spot problem makes network costlier.
2020 [126]	Authors proposed a 'Joint Density-Aware and Energy-Limited Path Construction algorithm for Data Collection, called DEDC algorithm'. DEDC outperforms the existing data collection algorithm in term of network lifetime and energy consumption.	DEDC comprises of path constriction phase and path adjustment phase. The methodology consists of partitioning the target area into equal sized grids and all grids have been connected to form a route for MS to visit each grid in such a way the total route length satisfies a length constraint. Thereafter, DEDC checks the balance of each grid by assuming each sensor of each grid as an individual cluster and merge the clusters one by one to form a hierarchical clustering tree based on distance factor. Next initiates path adjustment phase which does the appropriate length segment to each balanced and unbalanced grid depending upon the amount of data existing in each grid.	Designed for static WSNs. Target area is assumed to be rectangle in shape limits the applicability DEDC. MS follows a length restricted path to collect data from each grid, might adversely effect the buffer time.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2020 [127]	Authors proposed a 'Destination Oriented Routing Algorithm (DORA)' which balances the energy consumption, double the network lifetime and reduces end-end transmission delay as compared to PEGASIS. More suitable for strip-based WSNs.	DORA works to form multiple chains hierarchy by considering accurate distance between the nodes, forwarding direction and optimal cluster size. Optimal transmission distance, optimal cluster size and data forwarding direction to sink has been obtained by estimating energy depletion during intra-cluster and inter-cluster routing, results multiple transmission paths. Each chain length of DORA is shorter than the chain length of PEGASIS	Though it is the modification of PEGASIS but still is not suitable for mobile WSNs as whole routing topology needs to reorganize with the movement of a single node.
2020 [128]	Proposed a 'Multi-objective Cluster Head based Energy-aware Optimized Routing (MCH-EOR)' algorithm. This approach improves network lifetime with the optimal selection of CHs, reduces energy consumption and hot-spot problem. MCH-EOR is the most energy-efficient optimized routing based on fitness function.	A fitness function has been formulated considering the parameters like proximity, communication cost, residual energy and coverage to form unequal clusters and to select optimal CHs. The evolutionary Sailfish optimization (SFO) approach is used to find optimal path from CH to BS for data transmission. This quickly finds the optimal solution by selecting best path. An alternate path keeps on available if the nodes over the optimal path are on verge of death. This enhances fault-tolerant capacity of network.	Designed for static and homogenous power level of nodes.
2020 [129]	Proposed an 'energy-efficient mobility based cluster head selection (EEMCS)' scheme for Dynamic WSN. EEMCS is mainly focused on improved CH selection and adaptability of network with nodes mobility in energy efficient way. Throughput of EEMCS is far better than CRPD. This mechanism offers re-clustering only when it is essential. Moreover, it is scalable, saves more energy and prolong lifetime in Dynamic network.	EEMCS uses acknowledgment messages to ensure reliable data delivery and to avoid message repetition by using time counter. For CH selection, weightage of each node is calculated based on the parameters: residual energy of sensor nodes, mobility level, node degree and node's distance from the BS and then energy level of each node is checked to satisfy the minimum energy threshold criteria to be a CH. The node with the highest weightage become a CH. Each node is allowed to join one CH based on distance parameter at a time and two clusters cannot join each other. Intra-cluster routing is done using a single hop mechanism with TDMA schedule and inter-cluster communication uses multi-hop minimum cost transmission path for data delivery to BS. Nodes mobility and failures are detected by acknowledgement messages and handled by re-clustering if required.	Designed for homogenous network only. Inter-cluster routing has latency limitation. Number of data packets delivery decreases with increasing the network area.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2020 [130]	Proposal of 'cluster-based two-level routing method' which reduces number of total hops, energy consumption as well as increases packet delivery ratio and network lifetime.	In clustering phase, nodes calculate their own score for turning to CH, based on their distance to BS, residual energy and centrality. Node with the highest score advertises itself as a CH and the node that comes second in meeting the criteria advertises itself as Backup CH (BCH). Once the energy of CH is depleted, it is immediately replaced by BCH. This reduces extra calculations in new CH selection. In routing phase, each cluster is divided into four non-empty sections with respect to CH position in center. Intra-cluster routing requires nodes to transmit directly to CH only if it has the highest score otherwise node has to transmit highest scorer nodes which further transmit to CH. Inter-cluster routing involves layering of CHs according to the distance from sink as first layer CHs are only one-hop away from sink, second layer CHs are two-hop away from sink and likewise until all CHs are covered and formed layers. Routing of data are done from higher layer to lower layer of CHs. Lower layer CHs use single-hop to route to sink.	Designed for static homogenous network only. Node has to choose highest scorer node in intra-cluster routing causes unnecessary transmission delay. Random selection of CH or BCH when there are more than one choice for MNs, should be replaced by optimal selection in this work.
2020 [131]	Proposed an 'energy-aware cluster-based routing protocol (ECRP)' to reduce excessive overhead due to frequent CH renewal process. ECRP has linear message and time complexity. It reduces and balances energy consumption of SNs.	This method avoids unnecessary re-clustering by executing setup phase once in the first round and then network continues to function for rest of the rounds until it is out of service. ECRP supports topology changes with the help of acknowledgement messages and fault tolerant mechanism. In this, energy is the only factor for CH selection and MNs join CH on the basis of cost function made up of residual energy and distance parameters. Intra-cluster routing involves TDMA schedule and inter-cluster routing involves the concept of choosing appropriate relay nodes for data forwarding.	Designed for homogeneous and static network only. No control on number of CHs requirement.
2020 [132]	Proposed an energy-efficient clustering protocol for heterogenous WSN in which nodes have been classified as normal and power nodes based on energy level. This protocol maximizes network lifetime.	A probability model and threshold function is formulated depending on sensors energy for CH selection. Nodes satisfying the threshold criteria becomes CHs otherwise the process repeats again.	Designed for static WSN. There is no optimal path choice in inter-cluster routing for delay concern.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [133]	Proposal of 'Hierarchical Data Dissemination Strategy (HDDS)' for a large-scale dynamic WSN to optimize energy consumption with coverage preservation and fault tolerability of the network. This distributed algorithm forms a load-balanced routing path and diminishes the hot-spot problem by means of an unequal clustering strategy. It is more reliable, extend the coverage lifetime and supports the addition of new nodes or the partial movement of existing nodes. HDDS has linear Time complexity.	HDDS operates in five phases, 1. Bootstrap: Involves ngh discovery and acknowledgement of local information by nodes, 2. Clustering: Involves division of network area into non-overlapping clusters and selection of CH nodes based on intelligent waiting time which is computed using residual energy, coverage significance, distance and transmission energy. A node which results a low waiting time considered as an appropriate CH. Adaptability of network come up with the variation in transmission radius of CH nodes as closer the CH to the edge server smaller its radius for energy conservation purpose. MNs choose to join suitable CH by computing cost function having parameters, residual energy and transmission range. The CH resulting in higher value indicates a better association for MN. 3. Hop count calculation: CHs communicate with minimum hop counts, 4. Next Hop selection: CHs choose relay nodes based on residual energy, transmission energy and node degree. 5. Steady phase: It involves intra-cluster and inter-cluster routing, route recovery mechanism to handle fault at cluster or routing level.	Designed for homogeneous network. Authors should consider latency factor with coverage preservation and fault tolerance in clustering.
2021 [134]	Proposed a 'Tree-structured data aggregation' method. This approach significantly reduces energy consumption, provides more reliable and error free end-end communication.	The method consists of three phases. First phase is the binary tree formation in which nodes are organized in a sequential manner with the help of status given to nodes i.e. nodes are either visited or unvisited mode and get connected based on energy level. Second phase involves the process of secure data aggregation in which each parent has shared a key with all its children. Each node has to satisfy valid key allotment to the received data aggregation request for further processing. Third phase consists of reliable data aggregation process in which modified cyclic redundancy code (CRC) is utilized to find the error after every hop i.e. CRC code is computed on every bit of sensed data by each node and then data has been passed on to one-hop away node. If error is detected, request of retransmission is processed.	Process of retransmission causes data latency. Designed for static homogenous network only.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [135]	Proposal of a 'secure hybrid data aggregation method (SHSDA)' based on the combination of star and tree structures. SHSDA is reliable and fault tolerant with low overhead in route discovery. The stable path created between the source and the BS improves packet delivery ratio and reduces energy consumption.	In this method, whole network is divided into four sections and a star-like structure is formed in each section with the assignment of root node by BS on the basis of highest energy level and shortest distance from the center. Every child node is assumed to know its parent and grand parent based on the local information they received from BS and so each node selects the closest and highest energy node as its parent in each section with the sharing of a authentication key between every parent and child node. Then the phase of tree construction initiates connecting the selected root nodes to the BS which lies at root of tree. Both star and tree structures are made to be regular and fault-tolerant as when a parent node dies, child nodes are supposed to connect with the grandparent node with the new authentication key restoring the structure. For secure data gathering, leaf nodes send encrypted data to the parent node, parent node decrypts data, aggregates it with its own data and forward it further with the sharing of authentication key.	Designed for static homogenous network only. This method employs hop by hop data transmission may cause latency.
2021 [136]	proposed a 'multi-hop routing algorithm based on path tree (MHRA-PT)'. This algorithm optimizes network energy consumption and improves network lifetime and load balancing as compared to other tree related hierarchy routing.	Working of MHRA-PT comprises of CHs selection, working nodes assignment and transmission paths generation. Nodes closer to BS and having higher energy level than the defined threshold have been selected as CH nodes. Nodes except CHs, having residual energy more than average is chosen as working nodes and cluster partition happens by considering maximum sensing range of working node so that the whole cluster can be sensed by the working node. Data transmission happens by first sorting CHs based on distance to BS and CHs closer to BS called parent node. Each CH node except parent node, transmits aggregated data either via parent node or directly to BS by estimating residual energy and energy consumption of the relaying node and opting path respectively. This results multi-hop tree structure for routing.	Designed for static and homogenous WSNs. Only suitable for square-shaped network where sink is fixed at center.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [137]	put forth an energy efficient path planning strategy using multiple sinks. Author proposed stable election algorithm (SEA) which minimize message exchange, SP determination algorithm and multi-objective evolutionary algorithms (MOEAs) which optimally trace the route connecting all SPs. The proposed work is most suitable for Internet of things application of WSNs as it minimizes hot-stop problem, unbalanced energy consumption and improves network lifetime with connectivity.	The proposed work have been executed on three different types of network. First is homogenous WSNs, second is two-tier heterogeneous WSNs in which two types of nodes exist based on low and high energy level and third is three-tier heterogeneous network in which three different types of nodes exist based on low, medium and high energy level. The target region has been partitioned into equal area having same side length and no. of partition depends on number of MS. SEA algorithm initiates to form optimal number of clusters and selection of CHs based on node degree and residual energy, determine next CH rotation until the existing CHs' energy level gets lower than the defined threshold and data aggregation from CM by CH has been done in single hop. CHs transmit data to BS when BS traces path going through each SPs using MOEAs where MOEAs is objectifying path cost, travelling delay and distance.	Engagement of multiple mobile sinks makes the network costlier. Sensor nodes are assumed to be fixed after deployment.
2021 [138]	Authors proposed a 'Competition-Based Unequal Clustering Multihop Approach (CUCMA)' which maximizes network lifetime in large WSNs. This approach is distributed in nature and results optimal number of CHs and load balancing.	In CUCMA Nodes compete for CHs by calculating their own distance and residual energy and broadcast a message of being CH within a specified time threshold. Thereafter clustering happens on the basis of cluster radius, distance to BS, node degree and distances to around nodes. New CHs in other round are selected by involving one more factor i.e. distance to the existing CHs.	Assumption of target area is square in shape and nodes including BS are fixed restrict the application of the proposed work.
2021 [139]	Proposal of a 'novel non-uniform hierarchical clustering with dynamic route adjustment (NHCDRA) scheme using two MSs. This scheme is most suitable for a heterogenous WSNs. NHCDRA significantly reduces the data delivery delay and improves network lifetime when compared with state-of-the-art.	Heterogeneous network carrying two different types of sensors named as $\alpha-$sensor and $\beta-$sensor where the energy level of type$-\alpha <$ type$-\beta$. The target region has been partitioned into concentric circles of different radius and each ring region has been further divided into sectors by using straight lines which is based on communication range of type$-\beta$ sensor, results non-uniform clusters formation. CH is chosen in each sector based on highest residual energy and distance from center of cluster. Additionally there are boundary CHs (BCHs) lying over the boundary of each clusters. All the CHs accumulate and transmit data to BCH of adjacent cluster. Thereafter two MSs which are placed opposite to each other, move back and forth around the clusters boundary and collect the data from BCH if BCH is closer to MS else BCH relays its data to nearest node to pass on its data.	Designed for static network and engagement of two MSs makes the network costly. Assumption of obstacle free path between two sensor makes the work impractical. There is no defined path for MSs to visit BCHs.

Table 4 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [140]	Introduce 'divide-and-rule sectorization (DRS)' based hierarchical routing scheme. DRS approach brings balanced energy consumption of nodes and mitigate energy-hole problem, has better stability and network lifetime as compared to available state-of-the-arts.	The target area has been partitioned into 3 concentric circles where inner ring is depicted as I_C, middle ring as M_L, upper ring as M_U and outer region as O_R. Each ring is further partitioned into 8 sectors except I_C where 10 nodes are uniformly distributed in each sector. Nodes in I_O transmit their data to the forwarding node (FN) of M_U, FN further transmits to the FN of M_L, likewise data transmits to O_R and finally FNs of I_O transmit to BS. FNs are randomly selected and repeatation of FNs continue to happen until new FNs are going to exist. This balances load distribution.	This approach is suitable only for static and homogenous network where BS is placed at the center of target region and remain fixed after deployment.

$$X_{ij}(t) = X_{ij}(t-1) + V_{ij}(t-1) \tag{3}$$

where w is the inertia weight, c_1, c_2 are two non-negative constants, q_1, q_2 are two distinct randomly generated number in between [0, 1]. The update process is iteratively repeated until either an acceptable G_{best} is achieved or a maximum number of iterations is reached. A general mechanism of PSO technique is demonstrated in Fig. 8.

ACO techniques: ACO is the abbreviation of 'ant colony optimization'. This technique is introduced imitating the foraging behavior of ants seeking a path between their colony and source food. When ant moves in search of food, it leaves pheromone (a kind of liquid marker) along the path which evaporates over time. So other ants follow the same path by smelling the pheromone concentration. Moreover shortest path carries more pheromone quantity as more ants opt that path. With the movement of ants, quantity of pheromone updates everytime, this way a reliable and shortest path can be achieved using ACO technique. ACO has been used for a long time to solve route optimization problems in WSNs. Formulation of pheromone update function depending on distance factor, nodes residual energy and/or node degree has been done to discover optimal route/path in both static as well as mobile WSNs. A general mechanism of ACO technique is demonstrated in Fig. 9, taken from [141].

Following are some key points, we draw from Tables 5 and 6:

- **Optimal clustering by PSO technique:** PSO technique is having good performance in finding optimal location for CHs, RPs, SPs, anchor points (APs) and optimizing cluster size with optimal coverage. PSO routing techniques are based on to formulate one or more fitness function as first one is used to find optimal CHs as well as cluster size and second one is usually to find optimal data routing when there is a fixed BS or gateway node.
- **Optimal routing by ACO technique:** ACO technique is having best performance in bringing out optimal route between sensors with minimum energy cost and maximum network lifetime with the help of pheromone value. ACO technique gets upgraded with time like assignment of three different types of ants, one is used to find forward path, second is used to provide feedback and third is used to maintain the route. This promotes network connectivity and fault-tolerance. Also improved ACO techniques bring pheromone updation based on many factors like nodes residual energy, distance between the nodes, nodes degree, hop count, etc. This provides optimal next hop and enhances network life-cycle.
- **Hybrid of PSO and ACO technique:** Hybrid techniques of PSO and ACO, engage PSO to setup initial cluster topology and ACO to find the shortest route for data transmission and/or data aggregation. These tech-

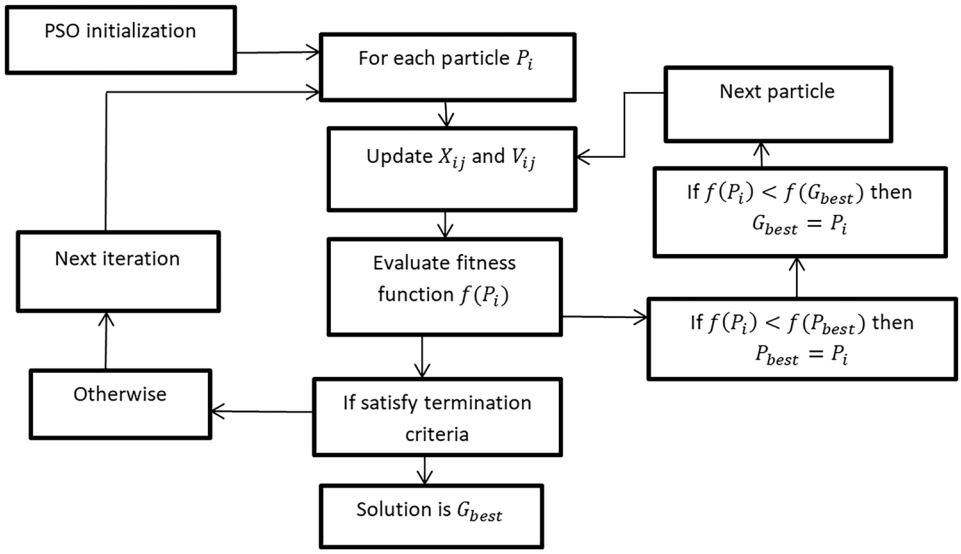


Fig. 8 General mechanism of PSO technique when the problem is of minimization

nique increase the convergence rate of ACO technique and results better network lifetime, reduced energy consumption and load balancing.

- **Engagement of MSs/SPs/RPs/APs:** Some PSO and ACO techniques engage MSs and SPs/RPs/APs for faster data aggregation and load balancing purpose. PSO-based technique utilizes fitness functions to find optimal number and position of SPs/RPs/APs and to obtain optimal trajectory for MS to visit each SPs/RPs/APs in an energy efficient and cost efficient way. ACO techniques update pheromone value along the route based on RPs load and distance between RPs and MSs.
- **Transmission delay and latency issue:** Both PSO and ACO handle contradicting multi-objectives in a problem and provide a balanced solution. Yet transmission delay/latency is the persisting issue in heterogenous and large WSNs in a cost efficient way. To reduce latency factor, some authors proposed solutions by engaging more than one MSs so that data aggregation and transmission time gets reduced. This reduces inter-cluster transmission delay only and makes the network expensive as well. Some ACO based routing techniques cause lesser end-to-end delay than PSO routing but in homogenous WSNs.
- **Mobility and robustness:** For routing in mobile WSNs, engagement of multi power level nodes are come into picture and CHs selection is done by formulating fitness function which consider nodes residual energy, nodes degree, nodes speed and distances to other nodes. Undoubtedly this makes the topology more robust and stable with lots of expense.

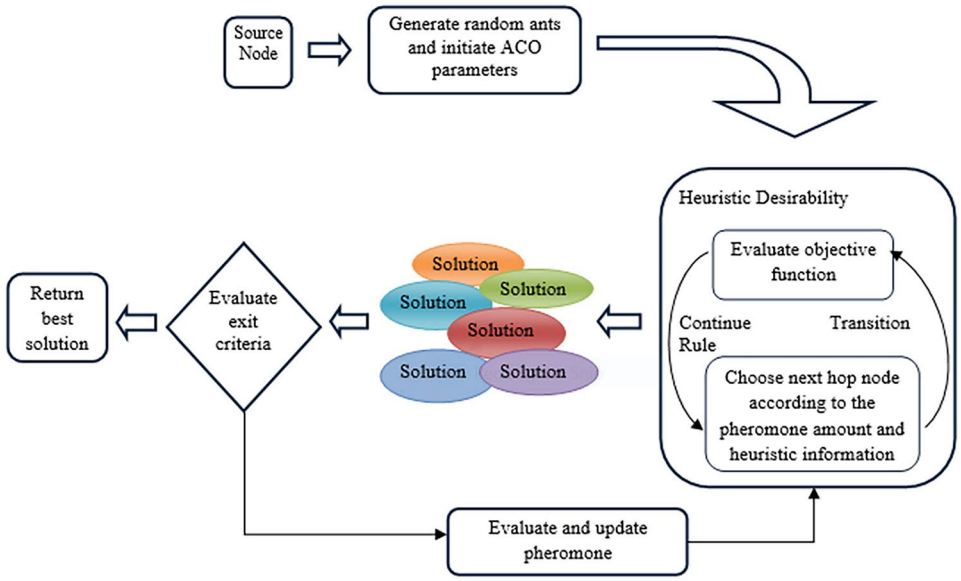


Fig. 9 General mechanism of ACO technique

Table 5 Overview of latest 'PSO based routing techniques' in WSNs

Reference	Contribution	Strategy/methodology	Limitation/future scope
2017 [142]	Proposal of a 'PSO-based energy-efficient routing and energy balanced clustering algorithms' for WSNs. The algorithms are able to cope-up with the failure of CH nodes. Algorithms save and balance energy consumption with prolonging network lifetime.	Authors formulated two multi-objective fitness functions with normalized parameters. One fitness function evaluate G_{best} route for energy efficient data transmission from gateway nodes to BS considering low energy consumption in transmission and nodes residual energy. Second fitness function brings out the optimal clustering by considering residual energy of each gateway nodes with even distribution of MNs to each gateway.	All nodes including gateway nodes are assumed to be static in the field. The proposed algorithm is centralized in nature and may cause routing delay since low energy path can be chosen, so not suitable for real-time application.
2018 [143]	Proposal of a 'PSO-based unequal and fault-tolerant clustering (PSO-UFC)' protocol. This protocol attempted to minimize energy-hole problem and load imbalance in intra-cluster and inter-cluster routing with connectivity assurance.	The target area partitioned into clusters of unequal sizes. Process of selection of master CHs initiated by BS, so a fitness function is formulated considering the parameters of highest residual energy, closeness to BS and lesser routing cost. After the clusters formation, PSO-UFC protocol constructs the multi-hop routing tree among the elected MCHs by formulating a fitness function with the parameters of residual energy of the next hop, distance of a nex hop from a MCH, distance of next hop from BS and node degree of next hop. Moreover, concept of surrogate CH (SCH) has been introduced as a backup of MCH incase of failure. MCH itself selects SCH in each cluster for optimal data routing.	Designed for static WSNs. There is lot of beacon message exchange among the nodes during the whole process, this causes overhead and energy expenditure as well.
2018 [144]	Proposal of 'a multi-objective particle swarm optimization (MOPSO)' algorithm which evaluates energy efficient path for MS. MOPSO successfully balances the trade-off between transmission distance and intermediate hop counts. MOPSO builds an effective approach of data aggregation through MS against hot spot problem.	MOPSO algorithm involves particle initialization using PSO technique and derivation of multi-objective fitness function for RPs based MS path design problem with the aim of minimizing both maximum average transmission distance and maximum number of total hop counts at any RP. Algorithm works to find RPs location and construct an efficient trajectory for MS connecting all RPs using the formulated fitness function. After maximum number of iterations, a global solution is obtained depicting the energy efficient path designs for MS where MS periodically travels and halts at each RP, for the purpose of gathering the sensory data.	Energy is not considered in examining objectives so some nodes may die prematurely.

Table 5 (continued)

Reference	Contribution	Strategy/methodology	Limitation/future scope
2019 [145]	Proposal of 'PSO-based uneven dynamic clustering multi-hop routing protocol (PUDCRP)'. This paper introduced a connecting line aided route construction method for energy efficient routing. Moreover, PUDCRP achieves balanced energy consumption of whole network, good scalability and reliability.	PUDCRP brings clusters formation as circles of distinct sizes based on distribution of nodes. Cluster size changes dynamically with the nodes count and distance from BS. Optimal no. of circles has been obtained by cyclic calculation and a multi objective function has been designed for optimal CHs selection considering nodes residual energy, no. of neighbor nodes and distances from nodes to the BS. Two fitness functions have been modelled to determine optimal positions of CH nodes and circular areas. first one is based on two important factors: 1.coverage rate 2.intersection over universal (IoU) and second depicts the absolute difference between actual number of nodes covered by the circular area and no. of ideal coverage nodes. If the distance between CH nodes and BS is less than some fixed threshold, CH nodes send data directly to the BS via single-hop routing otherwise a multi-hop route is established by the connecting line aided route construction method and the node with the smallest weight is selected as the next hop.	Designed for homogenous WSNs only. PUDCRP faces transmission delay in large-scale network.
2019 [146]	Introduced a special clustering method for heterogeneous WSNs targeting hole problem known as 'Energy Centers Searching using Particle Swarm Optimization (EC-PSO)'. EC-PSO shows good performance in terms of energy saving and improving network lifetime.	An improved PSO algorithm is used to find high energy centers so that nodes near the energy centers can be selected as CHs. EC-PSO provides optimal solution by designing fitness function considered with residual energy of those nodes which are two-hop away from the position of energy center. A low energy protection mechanism has been used to avoid weak nodes to become relay nodes by fixing threshold on residual energy. For inter-cluster communication, a mobile agent is utilized to gather the information from CHs which is placed at the center of the network initially. The collector always moves to the energy center with highest average energy.	EC-PSO is only applicable to the WSNs where nodes are evenly distributed and initiated with same power. Engagement of mobile agent makes the network costlier.

Table 5 (continued)

Reference	Contribution	Strategy/methodology	Limitation/future scope
2019 [95]	'A travel route planning schema with a mobile collector (TRP-MC)' to find a shortest route covering maximum no. of sensors, has been proposed. Energy consumption of TRP-MC is much less than LEACH. TRP-MC minimizes coverage overlapping with maximum coverage rate and improves data gathering efficiency of mobile collector. 'Sojourn points (SPs) are used to define stoppage location of mobile-collectors and 'anchor points (APs)' are used to denote SNs to estimate their distance from SPs.	TRP-MC divides the target field into regular hexagon with each hexagon is supposed to have a SP. Only one node need to keep active while other nodes go to sleeping mode momentarily in each division. PSO algorithm finds optimal position of SPs which cover maximum number of APs with minimum coverage overlapping. Authors formulate a fitness function as the ratio of overlapped coverage rate and total coverage rate and works to minimize it. Particles solutions are improved with iterations following P_{best} and G_{best} where speed and position of particles are restricted to some bounds so that particles don't move fast and go away from field. After max no. of iterations, G_{best} shows the optimal solution for SPs selection. After the SPs selection, ACO-based shortest path algorithm is applied to find shortest traveling route covering all SPs in single round.	Assumptions that nodes are static after deployment, sensor field is a rectangular area and a static sink is set at the corner of sensor field restrict the applicability of TRP-MC.
2019 [147]	Proposal of 'particle swarm optimization based selection (PSOBS) algorithm' which finds optimal and near-optimal RPs location for MS. PSOBS gains network lifetime, reduces energy consumption and packet loss rate.	PSOBS works in a centralized way as sink runs the PSO algorithm, specifies RPs and determination of routing path. A fitness function is formulated based on anticipated RPs number, maximum tour length of maximized tour and computation of path length through RPs. G_{best} provides optimal RPs.	PSOBS is limited to RPs selection only, lacking in optimizing clusters number and size. No optimal selection of CH nodes.
2019 [95]	'A travel route planning schema with a mobile collector (TRP-MC)' to find a shortest route covering maximum no. of sensors, has been proposed. Energy consumption of TRP-MC is much less than LEACH. TRP-MC minimizes coverage overlapping with maximum coverage rate and improves data gathering efficiency of mobile collector. 'Sojourn points (SPs) are used to define stoppage location of mobile-collectors and 'anchor points (APs)' are used to denote SNs to estimate their distance from SPs.	TRP-MC divides the target field into regular hexagon with each hexagon is supposed to have a SP. Only one node need to keep active while other nodes go to sleeping mode momentarily in each division. PSO algorithm finds optimal position of SPs which cover maximum number of APs with minimum coverage overlapping. Authors formulate a fitness function as the ratio of overlapped coverage rate and total coverage rate and works to minimize it. Particles solutions are improved with iterations following P_{best} and G_{best} where speed and position of particles are restricted to some bounds so that particles don't move fast and go away from field. After max no. of iterations, G_{best} shows the optimal solution for SPs selection. After the SPs selection, ACO-based shortest path algorithm is applied to find shortest traveling route covering all SPs in single round.	Assumptions that nodes are static after deployment, sensor field is a rectangular area and a static sink is set at the corner of sensor field restrict the applicability of TRP-MC.
2019 [147]	Proposal of 'particle swarm optimization based selection (PSOBS) algorithm' which finds optimal and near-optimal RPs location for MS. PSOBS gains network lifetime, reduces energy consumption and packet loss rate.	PSOBS works in a centralized way as sink runs the PSO algorithm, specifies RPs and determination of routing path. A fitness function is formulated based on anticipated RPs number, maximum tour length of maximized tour and computation of path length through RPs. G_{best} provides optimal RPs.	PSOBS is limited to RPs selection only, lacking in optimizing clusters number and size. No optimal selection of CH nodes.

Table 5 (continued)

Reference	Contribution	Strategy/methodology	Limitation/future scope
2019 [148]	Proposed a PSO-based routing algorithm to improve network lifetime. This approach reduces power consumption with improvement in load balancing.	A novel fitness function with three objectives has been designed by applying weighted sum approach to cope up with trade-off among the objectives. Objectives are are minimization of maximum distance between gateways and BS, minimize number of relay nodes for data forwarding and minimize relay load factor in the network. PSO-based routing provides an optimal routing path between each gateway to the BS following the proposed fitness function and dimension of each particle is equal to no. of gateway nodes.	In order to achieve load balancing, hop-count increases which causes transmission delay.
2020 [149]	Proposed a 'PSO-based dual sink mobility (PSODSM) technique' to reduce the energy expenditure of SNs in the heterogeneous WSNs. Use of two sinks make communication more energy efficient by reducing number of hops in large area networks, reduces overhead, transmission delay, improves stability and network lifetime.	Nodes with different energy level has been differentiated into normal, super and advance. Formation of fitness function for optimal CH selection, undertook factors like ratio of remaining energy to the initial energy, node degree, node centrality, distance between sink and node, CH count and energy consumption rate. Author placed two sinks at opposite corners and sinks moved towards each other to gather data from selected CHs against hot-spot problem. Also cluster MNs have been allotted with TDMA slots for sending data to the closest sink.	Assumption of target area is square shaped restricts the applicability of algorithm.
2020 [150]	Proposed a 'PSO-based Energy Efficient Clustering and Sink Mobility (PSO-ECSM)' algorithm for optimal CH selection and to optimize sink mobility such that hot-spot problem can be reduced in clustered based routing. PSO-ECMS decreases no. of dead nodes with respect to no. of rounds and prolongs network lifetime. Also throughput increases with load balancing.	Heterogeneity at nodes energy level differentiated nodes as normal, super and advance. Author formulated two fitness functions independently for the two optimizing purposes which has to be minimized comprising of multi-parameters that are nodes residual energy, distance factor between sink and other nodes, nodes degree and energy consumption rate. PSO-ECSM provides the optimal CHs and optimal inter cluster routing using fitness functions.	Assumptions of static nodes, target area is square shaped and sink has no energy constraint, restrict the applicability of the algorithm.
2020 [151]	Proposal of 'Energy-Balanced Cluster-Routing Protocol (EBCRP)' with five mutation operators in PSO algorithm to optimize number of clusters and cluster centroid. Reversing probability operator, Decreasing probability operator, Increasing probability operator, Swapping probability operator and Transforming probability operator are the five mutation operator which improve the performance of PSO algorithm by increasing search diversity and enhances the probability of finding better solution.	Any of the five mutation operators is randomly selected to apply to the position of i_th particle. Particles begin to search for the best clustering solution through iterations. These solutions converge towards the desired target by shorting the distance between nodes and their centroid while balancing the no. of nodes between clusters with the evaluation of P_{best} and G_{best} through fitness function. Also adoption of rotation of CHs based on the highest residual energy effectively balanced the energy consumption between nodes.	EBCRP algorithm dynamically selects CH for each cluster in each round, which reduces the stability of the network system. It is applicable to homogenous network only. This work can be extent to large sensing region where some relay nodes can be selected to assist the communication between CHs.

Table 5 (continued)

Reference	Contribution	Strategy/methodology	Limitation/future scope
2020 [152]	Introduced an 'Optimal Coverage Multi-Path Scheduling Scheme with Multiple Mobile Sinks (TSCR-M)' for large-scale WSNs. This enhanced network lifetime and Reliability.	Author introduces single hop trajectory scheduling i.e. sensors transmit their data directly to MS in just one hop. TSCR-M finds optimal parking slots where MSs stop for data collection subject to the constraints of maximum sensors are covered within the transmission disk of MS with minimum coverage overlapping. Author utilizes improved PSO algorithm for optimal search of parking slots where the fitness function evaluates the percentage of overlapped sensors.	Involvement of multiple MSs makes the topology quite expensive. Assumption that each sensor initialised with same power restricts TSCR-M applicability.
2020 [153]	Proposed a 'Mobile Double Cluster Head-Particle Swarm Optimization (MDCH-PSO)' algorithm to enhance the network lifetime and load balancing in hybrid WSN. By hybrid, it means SNs can be dynamic as well as static in nature. MDCH-PSO achieves good energy conservation of CHs, reliability and network longevity.	The author introduced the concept of female node (super power node), male node (CH) and child node (remaining MNs) in clustering. All nodes can move except female node and it is located at the centre of each cluster, do monitoring, clustering and data aggregation task by taking note of each child node ID, residual energy, one hop neighbors and distance to itself. This makes CHs live long in the network. Moreover male node selection i.e. CHs selection is done by PSO method by defining multi-objective fitness function considering residual energy, node density, mobile speed and distance to female node. Also female node predicts the mobility of child nodes using time stamp and so routing path is immediately updated on the movement of nodes.	Scalability will be the demerit due to high cost in deploying female node.
2021 [154]	Proposal of a PSO-based energy efficient clustering (PSO-EEC) protocol. PSO-EEC is designed for the heterogenous WSNs. The proposed work improves the network lifetime and stability as compared to MDCH-PSO, MOPSO and HSA-PSO protocol.	Authors engaged two types of sensor nodes, namely normal and power nodes where power nodes are having extra power than normal nodes. A fitness function has been formulated for the optimal selection of CHs based on the factors like energy ratio (i.e initial energy to residual energy), node degree, distance between nodes and CH. Further relay nodes are selected by employing fitness function considering the residual energy and distance factor of member nodes. Intra cluster routing uses TDMA schedule and in inter-cluster routing, CHs adopt multiple hops depending on the distance.	Designed for static WSNs only and multi-hop data transmission in large WSNs might cause latency.
2021 [155]	Developed a hybrid approach GAPSO-H where GA method is used to select optimal CH and PSO algorithm is used to trace optimal route for MS for data gathering or transmission. GAPSO-H is found to be computationally optimized and improves the network stability.	GAPSO-H method categorises nodes into normal, advance and super nodes on the basis of their power level. It undertakes hot-spot issue by integrating 5 major parameter in a single fitness function for optimal CH selection that are nodes residual energy, distance factor, average node energy, no. of surrounding neighbour nodes and energy consumption rate. Also for optimized sink mobility, three factors that are discovery of lowest energy CH, distance to CH and cluster length are considered.	Sink mobility makes network more costly. Also this work could extent to minimize end-end delay.

Table 5 (continued)

Reference	Contribution	Strategy/methodology	Limitation/future scope
2021 [156]	Proposal of a mobile data collector (MDC) based energy efficient clustering routing protocol for large WSNs. The focus of this approach is to reduce energy consumption and latency during transmission. Algorithm runs in polynomial time. Engaging MDC improves network lifetime and end-end data delivery.	The approach is the combination of improved variant of PSO i.e. variable dimensional-PSO (VD-PSO) and ACO methods. The VD-PSO method starts finding optimal data collection points (CPs) by iterating through variable dimensions so that each CP must surround with maximum CHs within a hop. So a fitness function has been formulated considering node's residual energy, neighbour centrality and coverage significance to choose optimal CP location using VD-PSO. After finding optimal CPs, ACO algorithm is used to form an optimal trajectory of the MDC to visit each CPs only once and get back to BS. Pheromone updation with time results an minimum length trajectory for MDC. Data aggregation starts with member nodes by CHs using TDMA and MDC uses CDMA to collect data from each CH by positing themselves at CPs.	Defined for static network. Involving MDC makes the network costly.

Moving to real-time routing protocols which are better helpful to minimize end-end delay and to obtain better QoS.

7.3 Real-time routing protocols

Real-time routing protocols are particularly concerned with delay-constraint or time-sensitive data transmission. These protocols are designed to minimize deadline miss ratio (DMR) and end-end data delivery with the maximization of throughput and reliability factor to support real-life wireless communication in a more efficient way [19, 72, 73]. A thorough review of well established real-time routing protocols from year 2002 to current, has been conducted and analysis of each technique has been documented in a tabular format where the analysis of some articles are done again and other are explored recently. Table 7 presents the objective/ contribution made by the proposed work, detail of methodology opted and associated limitations which might be overcome in future work. The definition of estimating end-to-end delay is distinct according to the proposed work like if the methodology is based on to determine packet velocity to travel, distance becomes the prime concern; some authors count number of intermediate nodes called as hops from source to destination; when methodology uses duty-cycle then total time spend during switching over sleep to awake, is considered for delay estimation. In Table 7, we have taken all the aspects to estimate delay where sometime delay is taken as propagation delay + transmission delay + queuing delay and so on. Real-time routing protocols are well adaptable to dynamic WSNs which makes them more applicable to real-life scenario. Most of them don't refer to establish topology and topology awareness like hierarchical routing. RAP [11], SPEED [12], MMSPEED [56] which are the very initial real-time routing protocols, determine velocity of data packets based on delivery deadline, urgency factor and distance from the destination node at the time of data transmission without establishing initial topology. However, their network lifetime doesn't last long as nodes die soon. After this, some authors [170–172] proposed delay constraint techniques which consider nodes residual energy as well while prioritising data delivery or to determine shortest route for data transmission. This reduces earlier node failure. Author [173] proposed a technique by involving location awareness and link quality to minimize routing error and packet loss at destination node. This maximizes throughput with reduced end-end delay. Some authors [8, 61, 174–176] proposed delay minimization techniques by opting cluster formation but these techniques are only suitable for static network. Technique in [177] engages MSs to collect data to minimize redundant waiting time, makes itself suitable for dynamic and delay-harsh network but by ignoring energy saving factor. Technique in [60] is non-topology aware protocol minimize end-end delay

Table 6 Overview of latest 'ACO based routing techniques' in WSNs

Reference	Contribution	Approach/methodology	Limitation/future scope
2017 [157]	Introduced an improved ACO based routing method which reduces energy consumption and extends network life cycle.	For optimal route evaluation, authors formulate a fitness function involving nodes residual energy and route length simultaneously with updation of pheromone concentration after each iteration. Larger the fitness value, more pheromone quantity gets update and the more optimal the route is. In this improved version, authors not only consider transmission distance of every next node but also distance between node and sink to save more energy. Also to reduce algorithmic complexity and energy consumption, authors limit the search space of ants by introducing the search direction angle which should be smaller than 90° between the line connecting a node and sink and the boundary line.	Assumption of a square field where nodes and sink become static after deployment limits the applicability of algorithm. Author ignores the effects of other energy levels (e.g., average energy, minimum energy). Extention to mobility and realtime constraints can be done in future work
2017 [158]	Proposed an 'ACO Energy Aware (ACOEA) routing algorithm' which offers multiple transmission path from source to sink. ACOEA, optimizes the ant colony algorithm in terms of angle and distance between nodes to obtain better multiple paths. This achieves balanced energy consumption and extended network life cycle.	In ACOEA, three types of ants are engaged, 1. Search path ants: search multiple energy efficient paths from source to sink by computing transition probability (having pheromone and distance parameters) at every next move, 2. Back-off ants: sent by sink node to source along the same path for updation of pheromone value and take record of nodes residual energy, 4. Repair ants: do the routing maintenance task and in case of node failure, path is repaired by replacing dead node with one of its suitable neighbour node. Authors introduce 'angle factor' so that any node which forms the largest angle between sink and source node taken as the next hop node. This helps in reduction of energy consumption and optimize the path performance. To search more paths, remaining links that are unused, selected successively to receive ants using the transition probability formula.	Authors assume the static model to design the algorithm. Dis-connectivity and hole problem can be the issue due to nodes replacement to repair path since these factors are ignored.
2018 [159]	Introduced a 'new data gathering method (ACO-MSPD) based on ACO technique' for non-uniform data producing sensors at different time period. ACO-MSPD finds the optimal no. of RPs and optimal trajectory (tour) for MS which needs to visit each RP to collect data within an imposed constraint of maximum tour length. ACO-MSPD minimizes average MS tour length, reduces the complexity to find suitable RPs. It adopts re-selection RPs dynamically during the change of weights of RP nodes. Therefore, it is not only minimizes the energy consumption but also balances the energy consumption among the RP nodes.	The proposed algorithm mainly consists of four phases, 1. Directed Spanning Tree (DST) formation: each SN chooses one forwarding node from its neighbour nodes of maximum load, 2. Optimal selection of RPs: RPs selection is done by improved ACO method where each ant chooses the next node as an RP for which the defined probability function (depending on pheromone value and distance) returns the highest value with the updation of pheromone value between the current node and the selected RP node, 3. MS path formation: an optimal tour by MS visiting each RPs to collect data using ACO technique and 4. Re-selection of RPs: If MS tour exceeds defined constraint length and/or when nodes generate non-uniform data then re-selection of few RPs takes place.	Complexity of Algorithm increases with increase in ant population and iterations. This method is designed for those network where sensors become static after deployment and a single MS has to engage.

Table 6 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2018 [160]	Proposal of a 'Hybrid ACO and PSO' based energy efficient clustering and tree-based routing protocol named as ACOPSO. This method prolongs network lifetime.	The algorithm first clusters the nodes based on their remaining energy. Then the hybrid ACOPSO method improves the inter-cluster data aggregation using pheromone factor which is deposited by ants moving from one CH to other. Involved ACO technique discovered a minimum cost spanning path between CHs and sink with the updation of pheromone after every iteration. Then PSO method initiates to optimize the cost of obtained structure further with the fitness function as minimum distance between the nodes.	Designed for static network where each sensor and sink remain static after deployment. Due to consideration of distance factor only, it may cause transmission delay when path of low energy is taken.
2019 [161]	proposed ACO based QoS aware energy balancing secure routing (QEBSR) algorithm. This algorithm improves network lifetime, end-end delay, reliability and load balancing.	QEBSR used improved models for calculating end-to-end delay based on lesser number of hope counts during transmission and trust worthy nodes has been chosen based on energy level and distance factors for data transmission to reduce packet dropping. While moving from source to sink, these ants store the information about the quality of nodes traveled such as energy consumption, residual energy, packets generated, packets received, packets transmitted and end-end delay and so update the pheromone level on their way back from destination to source. After maximum iteration, pheromone level keeps on updating and the path with maximum pheromone quantity becomes an optimal path for data transmission.	Designed for static and homogenous WSNs.
2019 [162]	Proposal of a novel 'Energy-Efficient Load Balancing Ant-based Routing Algorithm (EBAR)' for WSNs. This balances energy consumption, prolongs network lifetime and speed up the convergence of route discovery under the constraints of limited energy supply.	This new algorithm involves three components to achieve its aim: 1. An improved pheromone trail update method, based on energy levels of SNs and path length information, 2. An effective heuristic strategy which reinforces low energy cost routes in the route discovery phase based on expected cost metric and 3. An opportunistic broadcast scheme is adopted to control the packet overloading and congestion.	Assumptions of links are symmetric and sensors are isomorphic in functionality and capacity, restrict the applicability of the algorithm.
2019 [163]	Proposed an improved ant colony algorithm to optimize mobile WSNs in term of life cycle, imbalance energy consumption of sensor nodes and data gathering. Proposed method is having good dynamic adaptability and load balancing. A low power low error geographic routing based on ACO to enhance data reliability is seen.	Authors adopted both, global pheromone and local pheromone update strategies. To avoid the concentration of pheromone in the path that can be too high or too low, the pheromone concentration is set to range from some min to some max which in turn accelerate the convergence of algorithm. Updation of pheromone happens after every transmission along the path. A probability function has been defined to choose next move based on pheromone quantity. When the data needs to be transmitted, the optimization mechanism runs to calculate the current communication quality and distance of neighbour node.	Opted multi-hop communication routing may cause latency issue while maintaining load balancing and stability factors.

Table 6 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2019 [164]	Proposal of an 'ACO Optimized Self-Organized Tree-Based Energy Balance (AOSTEB)' algorithm in which an energy optimized intra-cluster route is selected using the ACO algorithm by MNs to communicate with nearest CH. AOSTEB algorithm involves rotation of CH when its energy becomes low, offering balanced load-distribution and helps in improving network lifetime.	The ant's movement creates multiple paths between MNs and their CH, optimal path is selected dynamically among the multiple paths on the basis of minimum energy consumption during transmission along the path. AOSTEB improves CH selection by introducing a parameter called as Probability of Election (PE) based on residual energy and neighbor nodes so a node which is having maximum PE value selected as a CH. Also selected CHs keep minimum distance between each other so as to have a balanced distribution of clusters inside the network. Data transmission are done when CHs form a tree-like structure in which only one CH on behalf of all CHs communicates with the BS whose residual energy is greater among all.	AOSTEB is designed for static network only. Initial random route selection causes redundant data transmission. Moreover, only one CH transmits data to BS directly, it may cause congestion at CH node as well as transmission delay.
2020 [165]	Proposed a 'distributed data gathering technique (GTAC-DG)' where the decision making skill of game theory are used to find optimal set of RPs and enhanced ACO technique is used to find optimized route among the RPs for MS visit. GTAC-DG helps to reduce redundant data transmission, energy consumption and delay in data delivery. A reliable and intelligent trajectory of MS has been resulted, extending the lifetime and energy conservation. As selected RPs directly transfer data to MS, this leads to better balance in energy consumption and reduces communication overhead.	The technique starts with game theory Based RP-selection algorithm where nodes involve in a local-game with its neighbors competing with each other to be a RP depending upon the relative energy and load of player nodes. Larger the relative energy and load, more is the probability of being a RP. After the suitable RPs selection, trajectory design phase is executed i.e. travel path for MS visiting all RPs, same as TSP problem. So a modified ACO method is proposed by authors, formulating a tradeoff function between considered parameters that are load factor of RPs, pheromone value and distance between RPs and MS to have an optimized path. Also RPs with higher load are prioritized first to visit by MS to reduce retaining data to buffer for a longer period of time.	Assumptions of sensors are immobile and have same communication range restrict the applicability of algorithm.

Table 6 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [166]	Proposal of 'Multi-sink placement and any-cast routing (MPAR)' and 'Extended multi-sink placement and any-cast routing (EMPAR)' involving two level architecture where lower level comprises of all cluster groups and upper level comprises of sink nodes and their related activities. With the increase of clusters numbers and decrease in cluster size, algorithms can balance nodes energy consumption, reduce transmit power consumption and consequently prolong network lifetime.	Methodology of the proposed algorithms is as follows, 1.Sensors cluster themselves based on Euclidean distance where MPAR selects CHs whose energy level is maximum while EMPAR considers both residual energy and distance factors along with the history of sensors selected as CHs, 2.Formation of shortest path tree at lower level through which each sensor transmits its data to the corresponding CH, 3.This engages improved PSO algorithm to choose optimal locations for multiple sinks to place where a fitness function considers distance between the CHs and sinks with data length of each cluster and 4.Construction of high level any cast routing tree connecting all the CHs with multiple sinks for low cost data transmission has been done by improved ACO algorithm using pheromone based probability function. The employed probability function jointly considers residual energy and Euclidean distance for selecting next hop CH.	With the increment of sensors, time complexity increases. Involvement of multiple sinks makes network costly. As the number of sinks increases, finding the best location for the sinks becomes more complicated.
2021 [167]	Proposed an energy-constrained and delay-sensitive data aggregation approach. This approach gives shortest possible MS traversal route in polynomial time. Latency is minimized by employing one hop communication between CHs and MS and shortest path traversal of MS.	This method finds SPs such that tradeoff between energy preservation and data transmission delay gets balanced. This formulates the optimization problem of minimizing no. of SPs with one-hop connectivity to CHs. Since more SPs causes delay and less SPs causes energy hole issue so this approach aims to find optimal SPs and shortest path visiting each SPs by MS. SPs nomination is based on density factor i.e. density of nodes around CHs and coverage overlapping. An ACO method has been employed to find shortest travelling path for MS, using transition probability function which is further based on pheromone concentration and distance. CHs use TDMA schedule for data transmission among themselves.	Applicable where nodes remain static. Obtained traversal route for MS can be less reliable since nodes residual energy has been ignored.

Table 6 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [168]	Proposal of 'Energy-efficient cluster-based routing protocol using unequal clustering and improved ACO techniques (ECRP-UCA)'. ECRP-UCA is able to achieve global optimal solution and prevent quick convergence to local solution. Moreover it treats connectivity, fault tolerant issues of large network, balances energy consumption and improves network lifetime.	This method uses a new parameter i.e. no. of backward relay nodes in previous round with distance from the sink, residual energy and node degree to determine cluster radius. This measure gives unequal cluster formation. MN selects its CH according to the multi-objective cost function which is based on residual energy and distance parameters. Also to manage connectivity in large-scale network, ECRP-UCA refers higher energy nodes as relay nodes for ensuring multi-hop connectivity between isolated nodes and distant CH nodes. Strategy like replacement of dead nodes by already backup nodes without reconfigure whole network is adopted to ensure fault-tolerant network. For energy efficient inter-cluster routing, author formulates improved ACO (IACO) method with an addition of important parameter called NPPRN (No. of preferred probable relay nodes) with the energy of next hop node and distance from the current node to the next node.	Designed for static homogenous network only. Computational time increases with the increment in the value of parameters.
2021 [169]	Proposal of an extended ACO-based mobile sink path construction for event-driven WSNs, abbreviated as ACO-MSPD. This helps to improve network lifetime, hot-spot issue, load balancing and throughput. The present work outperform its previous version and EAPC.	The first task is to construct a connected directed spanning tree (DST) of the given network. Thereafter ACO is used to generate the RPs location by updating the ant's pheromone quantity iteratively. A probability function has been constructed to trace delay-aware path for MS to collect data from each RPs and MS prioritized RPs based on available data size. Concept of virtual RPs (VRPs) is the new approach used in this work. VRPs are those SNs which are more closer to MS than the RPs so they directly transfer their data to MS when comes under its communication range. This reduces the buffer time, redundant transmission and delay as well.	Nodes except sink node are supposed to be static in nature after deployment. So the proposed work is not suitable in mobile WSNs. Also the preferred DST model of this work is less fault-tolerant.

Table 7 Overview of 'Real-time routing protocols' in WSNs

Reference	Contribution	Approach/methodology	Limitation/future scope
2002 [11]	Proposal of a new packet scheduling policy i.e. 'RAP' which minimizes packet miss ratio within a deadline. RAP is most suitable for large-scale WSN as it minimizes communication overhead of large network.	In RAP, transmission of data packets at a node are prioritised according to the requested velocity. This velocity assignment depends on the distance of node from the destination and deadline given to the required data packets to reach to its destination. Shorter the deadline and larger the distance more is the velocity requirement. So packets with higher requested velocity are prioritised first. Author investigates two types of velocity assignment: Static velocity monotonic (SVM) and Dynamic velocity monotonic (DVM). SVM, velocity is calculated once at the initial node and remain the same while DVM, velocity may vary at the intermediate nodes accordingly.	RAP doesn't consider residual nodes energy in prioritising and determining velocity as it may cause earlier death of nodes and hence detruncate network lifetime.
2003 [12]	Proposal of 'SPEED' protocol which reduces end-to-end delay, congestion and voids. SPEED helps to maintains desired speed of data packets in transmission. It supports feedback control and saves energy by reducing number of control packets.	SPEED protocol consists of beacon exchange scheme, delay estimation scheme, stateless non-deterministic geographic forwarding algorithm (SNGF), neighbour feedback loop (NFL) and back-pressure re-routing. SNGF used by SPEED to select highest speed node as relay nodes, calculation of relay speed is based on distance of ngh nodes and local information exchanged by beacon messages. Delay estimation between the sender and receiver nodes is done with the exchange of acknowledgement messages. NFL is used to maintain single hop relay speed with the knowledge of nodes miss rate. Back-pressure re-routing is utilized to reduce the congestion inside the network while transmission. Each node keeps the detail of its neighbour nodes and calculates packet progress speed of each neighbour node to forward its data to the node whose progress speed is greater than the predefined lowest speed limit.	If a node cannot find a neighbour node with progress speed more than predefined speed, packet will dropout, hence less reliability. This protocol ignores energy metric in speed determination and routing.
2004 [183]	Proposal of an 'energy-aware delay constraint routing algorithm' for both static and dynamic sink engagement. This approach provides the least cost path meeting end to end delay constraint, reduces average delay and packet miss ratio. Also this limits unnecessary re-routing which helps to save energy.	Extended Dijkstra's algorithm is used to find k disjoint paths in ascending order of cost where the cost function for each link is based on communication distance, nodes residual energy and transmission error rate. Appropriate transmission range for each node is calculated, satisfying time deadline. The method is extended to support sink mobility in the network by employing automatic re-routing approach. If the MS lost connectivity due to its movement, new nodes are directed to connect with the sink to relay packets and routes are updated by adding new hops. If the sink moves close to the source nodes, it over-hears the current transmitting data traffic, reduces unnecessary relaying and end to end delay. Re-routing only happens when already established routes become less energy efficient.	This does not support scalability factor. Utilization of class-based, priority queueing mechanism rises complexity and makes the network costlier for the resource-constrained sensors.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2006 [56]	Proposal of a 'Multi-Path and Multi-Speed (MMSPEED) protocol' considering timeliness and reliability factor. MMSPEED achieves guaranteed end-to-end delay and reliability requirements in a localized way. Suitable for more scalable and large dynamic WSNs.	MMSPEED replicates the SPEED protocol in designing multiple SPEED layers. This protocol works on two important notions: virtual isolation among the SPEED layer and dynamic compensation of local decisions. Virtual isolation aims to classify incoming packets into an appropriate speed layer based on their speed, this reduces transmission delay of lower speed packets by differentiating them from higher speed packets from the queue. The dynamic compensation is the adjustment in speed at intermediate nodes while transmission to meet end-to-end deadline. Delay is estimated at each node after every hop using elapsed time concept for updation of speed if necessary. To enhance the reliability level, MMSPEED adopts multiple path data forwarding. Number of forwarding path is depending on packet reliability level as low reliability packet is delivered using a single path while a high reliability packet is delivered via multiple paths.	MMSPEED requires more complex calculations for multiple path selection and dynamic compensation, it may lead to more power consumption. This is designed for short-lived network only.
2006 [170]	Proposal of an 'energy-aware dual path real-time routing scheme' for delay sensitive data packets. This approach focused on achieving reliability and improving network lifetime. It also reduces congestion at intermediate nodes by minimizing redundant transmission.	Packets are prioritised to deliver on the basis of urgency factor depending on path distance and packet deadline. More is the urgency factor is for a packet at a node, earlier the packet has to deliver. Also on the basis of urgency factor, the value of distance traveled from the source node towards the destination node is decided. Intermediate nodes are prioritised using the calculation of energy and delay so that packets can reach final destination on time. This scheme involves dual transmission of data packets from the source node i.e. source transmits same data packets to first priority node as well as second priority node in its surrounding and after that only one path is taken to route data upto final destination. Additionally, Author engaged MAC layer to minimize delay on its part by stopping post backoff for real-time data packets.	This approach ignores nodes energy consumption to achieve QoS and delay requirement.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2007 [171]	Proposal of 'LR-ENR and HR-ENR' real-time routing algorithms with the aim to have energy efficient delay constraint routing with achieving optimal lifetime in the connected network.	At first, author implement a LPP objectifying maximum of minimum network lifetime avoiding delay factor for energy efficient routing. A distributed algorithm is proposed which iteratively solve this LPP by giving least cost path routing where the cost of each path is either the sum or the maximum of the cost of the nodes on that path. Due to ignorance of delay factor, author extends the approach further by incorporating delay guarantee which has been named as Level restricted energy efficient routing (LR-ENR) and Hop restricted energy efficient routing (HR-ENR). LR-ENR limits the number of iterations and HR-ENR limits the length of path by minimizing number of hop counts to meet the deadline.	It does not consider link's reliability. An unreliable link in any route may cause longer delay because of higher retransmission probability, which in turn increases energy consumption.
2008 [173]	Proposal of a 'real-time routing protocol for load distribution (RTLD)'. RTLD distributes task of data forwarding, hence prolongs network lifetime. RTLD reports high performance in term of delivery ratio, control packet overhead and power consumption.	This method is location aware and uses geo-cast with directional forwarding to engage multiple paths for data delivery. Routing consists of forward metrics calculation, forwarding mechanism and routing problem handler. Packet reception rate (PRR), Packet velocity and residual energy are forwarding metrics, defined to calculate optimal next node and data forwarding is done using either unicast or geo-directional cast routing depending on the direction and distance of receiver node and its surrounding. Average delay in data routing after every hop is calculated to ensure packet delivery before the deadline, this increases throughput. When there is no forwarding node found to route data to destination then routing problem handler is used which has two recovery method: fast recovery, it can be done when a node's transmission range covers the whole hole and slow recovery is done by using feedback control to acknowledge backward nodes about the hole.	Designed for static network only. Inspection of link quality based on PRR is time taking and leads to energy consumption.
2008 [184]	Proposed a new routing algorithm undertaking energy, delay and reliability factors simultaneously. This increases sensors lifetime and works for low mobile sensors also.	Approach is the extension of AODV protocol where nodes discover routes in request-response cycles with the difference of opting energy efficient paths on which transmission delay is lesser than the allowed delay.	Delay is measured on the basis of hop counts only. Interference and distance parameters should be taken into account for transmission delay.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2009 [58]	Proposed a new real-time routing based on two-hop neighbourhood information named as 'THVR'. THVR eliminates topology hole problem while routing and reduces DMR. Achieved a better tradeoff between load-balancing and transmission delay.	The approach of calculating and requesting required velocity for every incoming packet at each forwarding node is similar to SPEED protocol but with the difference of considering two-hop away nodes from the source node for next move. Also in THVR, a joint metric of required potential velocity and nodes residual energy is considered to determine next forwarding node. Delay estimation at each forwarding node involves MAC delay and transmission delay is done to predict packets deadline. Initiative Drop Control method is conducted when there is no two-hop forwarding node available with the required speed. Then dropping of packets with the request of retransmission takes place.	Data reliability is low and with the dropping of packets, chances of overhead and energy consumption increases.
2012 [178]	Proposed a 'Fast Time-Dependent Shortest Path (FTSP) Algorithm' for end-end delay reduction in WSN. FTSP algorithm achieved least latency. Moreover, another version of FTSP is also proposed, named as 'FTSP-M' to deal with mobility of sensors. Both the proposed algorithms have polynomial message and time complexity. FTSP-M runs concurrently at multiple nodes if there are multiple node updates.	FTSP algorithm works similar to Bellman-Ford algorithm with the inclusion of β-synchronizer to bound the message complexity for time-varying network. FTSP consists of two major steps: first is construction of spanning tree and second is computation of shortest path from every node to sink, layer-wise i.e. nodes which are lying in the nearest layer to sink are computed first and so on. Control messages play the task of doing acknowledgements about distance between nodes. This algorithm is further extended to dynamic networking named as FTSP-M. FTSP-M updates shortest path only for updated nodes without disturbing whole network and the process of spanning tree formation accomplished with the help of control messages.	This cannot solve the void path and routing loop problem which affect the real-time communication.
2013 [60]	Proposed an 'enhanced real-time routing protocol with load distribution for mobile WSNs (ERTLD)', extension of RTLD. ERTLD achieves higher throughput and reduces end-end delay in dynamic WSN. Also ERTLD solves hole problem and saves calculation time by utilizing RSSI as optimal forwarding metric.	This method engages corona mechanism instead of geocast-directional routing. Corona mechanism involves the formation of concentric circles centred at sink, more the distance of SNs from the sink, farther they lie in the concentric circles having more radius. Corona level increases parallely with the radius of concentric circles and so data travels from higher corona level to lower to finally reach the sink. Optimal forwarding parameters like RSSI (received signal strength indicator), delay per hop and residual energy are simultaneously considered to make choice for next move. Backup corona mechanism is used against network hole while data forwarding. In this, a backward control packet is send to stop transferring data on the same route and re-routing is done.	This method does'nt consider lifetime and energy saving factor explicitly.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2014 [179]	Proposed two distributed query-driven data forwarding approach for delay sensitive network and bring out good tradeoff between energy efficiency and congestion avoidance. Development of 'First Forwarding Nodes (FFNs)' and 'Second Forwarding Nodes (SFNs)' enhance network connectivity and mitigate void path problem in real-time communication. Proposed methods can be applicable to mobile network also and establish an energy balanced delay sensitive routing.	The concept of FFNs and SFNs are introduced so that each node first uses FFNs as relay nodes until FFNs go out of energy, afterward SFNs work as relay nodes. Flooding of query messages by BS makes every node to get their FFNs and SFNs on the basis of hop counts. Data routing is done by choosing minimum cost link which is depending on residual energy and channel congestion degree in such a way that it constitutes a directed acyclic graph avoiding looping of data transmission. Proposed first algorithm utilizes routing messages for every transmission causes energy consumption so the modified algorithm is proposed to reduce this by employing the prediction of link quality and local information based on previous round.	Assumption of communication link is error free makes this method quite impractical. The problem of nodes isolation and voidness persists in sparse network.
2016 [174]	Proposal of a 'delay-constrained energy multi-hop (DCEM)' algorithm to find optimal solution of TED (Tradeoff between Energy and Delay) optimization problem in real-time cluster-based routing.	A new energy-cost function is designed with the delay estimation including link, queuing, transmission and propagation delay to find the least cost route satisfying delay constraint. DCEM does network organization task including clusters, route formation and data transmission i.e. multi-hop inter cluster routing. DCEM operates the same task after every round. Two parameters α, related to CHs' residual energy and β, related to distance between CHs and sink, are engaged to establish the balance between energy consumption and end-end delay.	Designed for static network only. Consecutive network organization causes energy consumption.
2016 [59]	Proposal of an 'energy-efficient adaptive routing protocol (EE-ARP)' for real-time communication. EE-ARP ensures energy efficiency, better routing performance and increases network lifetime with least latency and few packets dropping.	This work involves the identification of slow data packets by calculating expected delay whose transmission deadline to reach destination are expired at any intermediate node and removal of those expired or much delayed packets from the queue of that node to eliminate its transmission. In this way, a queue maintains data packets of sufficient residual deadline only. Also nodes can vary their transmission power based on distance using two-hop ngh information to reduce dropping of data packets and retransmission.	Designed for homogenous, static and dense network only. Dis-connectivity and voidness can be a problem in less dense network.
2017 [172]	Proposed a 'Delay-constraint energy efficient routing protocol based on Lagrange relaxation (DELAR)'. This heuristic algorithm has polynomial time complexity. Reduces average delay and energy consumption. This method has been extended to find multiple shortest path.	DELAR consists of two phases: cluster setup and routing phase. An optimization problem is formulated in which link cost is taken as an aggregated function of energy consumption and transmission delay to find the least cost path from the set of all paths from SNs to sink satisfying the delay constraint. Two heuristic algorithms are proposed, algorithm 1 is to find the optimal multiplier α which belongs to some interval $[a, b]$ which is further engaged in algorithm 2 to find optimal path of least cost satisfying end-end delay.	Analysis complexity increases with the increment of number of SNs. Defined for static network only.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2017 [7]	Proposes an energy efficient link delay aware routing algorithm. This approach is very well suited for real-time wireless communication as it reduces end-end delay and balances the tradeoff between energy consumption and delay factor.	The routing mechanism comprises of four phases, i.e., neighbour discovery, metric calculation, score calculation and node selection. Exchange of 'Hello' msg tracks the neighbour list. Residual energy, end-end delay and transmission counts are considered metrics and PRD value is the required score based on the given metrics. PRD demonstrates the predicted remaining counts of successful deliveries within one unit of delay. Next relay node has been selected for routing using PRD results best link quality, shortest end-to-end delay and highest residual energy.	Designed for directed and homogenous network. Authors did't provide the complexity of the proposed work with the change of network density.
2018 [182]	Proposed a distributed 'Minimum-Delay Energy-efficient flooding Tree (MDET)' algorithm to find the optimal solution to the multi-objective problem of minimizing delay and energy consumption in flooding based data transmission. MDET shows a very good balance between flooding delay and energy efficiency with minimization of flooding cost and delay.	MDET is designed to bring out the balance to energy-delay tradeoff in low duty cycled network with unreliable links by constructing energy-optimal tree with bounding flooding delay. It involves the process of Expected Transmission Count (ETX)-based shortest path formulation and corresponding nodes enhance energy efficiency locally by choosing its parent node with best link quality and loop free transmission. A biased back-off mechanism is adopted in flooding tree construction to reduce control message overhead.	Designed for static network with low duty cycled and unreliable wireless link. Reduction of delay is usually achieved at the price of detruncate network lifetime.
2019 [177]	Proposed a shortest path strategy for MS for data aggregation while ensuring connectivity in a typically connected/disconnected network. This approach reduces end-end delay with load balancing. The proposed work is most suitable for delay-harsh applications where the delivery delay is much more important than energy efficiency.	The methodology consists of two phases: first phase determines the anchor points (APs) where MS visits to collect data. APs are selected such that some are close to network center and some are far from network center. So the target area has been partitioned into grids having APs and joining of 4 grids having distant APs from center results quadrilateral like structure. Second phase adds new APs to the existing structure until there exists an AP in every segment. The new APs must be near the current path as much as possible. DAAP considers disconnected segments of the network in an decreasing order of their minimum joining cost to obtain a final and shortest path for MS.	The proposed solution is localized and approximate one. This approach is more inclined to minimize delay so energy consumption is still persisting issue.
2019 [8]	Proposed a 'delay aware energy efficient reliable routing (DA-EERR)' technique for a large scale heterogeneous WSNs. The proposed work deals with hot-spot problem, improves network lifetime and reduces packet loss. This technique achieves a good balance between energy efficiency and fast data delivery in the large densely deployed heterogeneous WSN having mobile sink.	In DA-EERR, the network area has been partitioned into multiple rings with three types of nodes based on energy level i.e. energy of type A > type B > type C. Type B and C nodes sense the data packets and type C nodes collect and forward data while carrying the information of mobile sink node. For quick and easy data transmission, search space towards the destination node has been reduced. Every source node chooses relay node of highest energy level within minimum distance to transmit its data to mobile sink.	Defined for static network. The proposed work depends on excess beacon messages transmission.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2019 [61]	Proposal of a new 'Delay Constraint Energy-efficient Multi-hop routing algorithm (DCEMRA)' involving a new rule-based clustering and routing which builds balanced intelligent clusters and low cost routes. The major advantage of DCEMRA algorithm is the improvement in reliability, packet delivery ratio, network lifetime and reduction in energy consumption and delay.	Clustering process consists of three major steps: CH selection, cluster formation and tradeoff of energy and delay (TED) calculation. Selection of CH is based on nodes energy, RSSI and agent computed value to acknowledge ngh. Updation of CH need to be done when it doesn't able to satisfy energy threshold. TED factor adjusts mobility parameters over the CHs energy expense and distance of CHs from MNs to bring out the balance between energy consumption and end-end delay. DFS algorithm is utilized to find low cost path satisfying the delay constraint during inter-cluster routing. The whole routing process consists of link and route cost calculation of all possible paths, end-end delay calculation of selected low cost shortest paths and multi-hop data communication.	Designed for static homogenous network and work in a centralized way by sink node which is located at the centre of the field. Engagement of intelligent agents makes the process costly.
2020 [175]	Proposed a 'Delay-aware Green Routing Protocol (DGRP)' to deal with hot-spot problem in WSNs by engaging mobile sink node. Engagement of mobile sink nodes reduces the energy hole problem, end-end packet delivery, energy consumption and outperform general ring routing protocol (RRP) in large WSNs.	DGRP partitions the target area into optimal number of rings, by considering the length of the sensor field and communication range of the sensor nodes. The nodes which are closer to sink node in every ring chosen as high tier nodes which keep the location information of sink nodes and then high tier nodes, belonging to the same partition, reshape the ring by joining themselves. Second type of nodes are called low tier nodes which pass their data in one hop to high tier node belonging to the same ring. Thereafter high tier nodes pass the information either directly to sink node or minimum length neighbour high tier node by setting an acute angle towards the direction of sink node.	Homogeneous and static sensor nodes having same radio range limits the application of the proposed work.
2020 [181]	Proposed a 'transmission delay minimization based on adjustable duty cycle (DMADC)' routing scheme. DMADC makes better use of node energy and reduces network latency than previous methods. DMADC scheme can reduce the end-to-end data transmission delay by $10.25\% - 26.37\%$ while maintaining the network lifetime and increase the energy utilization rate by more than 25%. The DMADC scheme improves network lifetime by about 30% while maintaining end-to-end latency.	DMADC is Basketball Net topology (BNT) based approach and each node is assumed to have two state: sleep and awake. DMADC comprises of two algorithms; Algorithm 1 computes the shortest path by computing minimum transmission delay of each relay node from source to sink node i.e. source node use a relay node of minimum transmission delay where transmission delay is the sum of sleeping delays along the path. Algorithm 2 is the duty-cycle adjustment approach. When there is no node in the active state to forward data, parent node informs the child node to adjust its duty-cycle from sleep to awake. Therefore nodes which are lying far from sink node are kept on higher duty cycle and nodes near the sink node are kept on low duty cycle. This helps in reduction of energy consumption of nearby nodes and avoidance of hot-spot problem.	Designed for static and homogenous network, makes the work quite impractical.

Table 7 (continued)

Reference	Contribution	Approach/methodology	Limitation/future scope
2021 [176]	Proposed a 'Delay-Sensitive, Reliable, Energy-Efficient, Adaptive and Mobility-Aware (DREAM)' routing protocol, incorporating nodes mobility and network heterogeneity. DREAM is the cluster-based solution to the multi-objective problem which is simultaneously dealing with energy consumption, network lifetime, latency, throughput, reliability and scalability.	In DREAM, target area has been partitioned into sectors where nodes are randomly spread and BS at the center. From each sector, a CH is selected based on the highest power among the competing nodes which are eligible to be CHs on the basis of residual energy and distance factor. MNs from each sector send their data to the CH and CH, depending on its distance from BS, can transmit data directly to BS or use other CH as a relay node to pass the data to BS.	This work utilizes single and two-hop routing in 2-D model of WSNs persisting redundant transmission issue. The work can be extended to 3-D model incorporating multi-hop routing and data redundancy issues.
2021 [180]	Proposed a real-time application based routing scheme to minimize end-end delay using two-hop information. This approach establish a good balance between energy efficiency and QoS level.	Every node keeps the record of its hop count from BS as height i.e. node with one hop away from BS has height 1 and so on. Data transmission has been carried out by estimating every two-hop away node' s residual energy, delay and link quality and so a PRI (Potential relay information) function has been formulated to estimate the possible forwarding nodes based on these parameters. After computation of PRI value, each node computes its neighborhood state index (NSI) based on PRI which depicts the quality of its surrounding nodes. Higher the value of NSI, higher the chance of being selected as a relay node during data transmission.	Designed for static and homogenous network. The proposed scheme utilizes two-hop information at the expense of computational complexity.

by considering RSSI (received signal strength indicator), delay per hop and residual energy simultaneously to make next move. [7] considers more factors like packet delivery, nodes energy, nodes degree, distance, retransmission etc to determine next routing hop to balance energy consumption with delay minimization. Authors in [58, 179, 180] consider hop counts to determine minimum delay path while authors in [181, 182] estimate end-end delay based on sleep/wakeup delay. After going through each technique of Table 7, we draw some conclusion:

- Prioritizing end-end delay minimization effects the nodes energy adversely in such a way that they die soon in the network so saving nodes energy should be taken equally important. This improves overall network performance.
- Topology aware real-time routing techniques balance end-end delay and energy consumption in an optimal way but they are lacking their performance in dynamic network. Also some are only designed for homogenous network.
- Indeed proposed techniques in this field enhance packet delivery ratio, improve throughput, reliability, reduce DMR and end-end delay at the cost of more energy consumption than usual or by making network costly by opting MSs or backup nodes.

8 Discussion and comparative study

In this section, we discuss about prime benefits, persisting issues and future scope of each of the reviewed topology control and routing technique along with the comparison between the existing solutions.

8.1 Topology control techniques

Graph based range assignment and power control Among the graph based approaches which are mainly focussed on range assignment problem and transmit power control, some of the work [15, 84] has adopted arbitrary and fixed transmit power assignment to nodes and some [17, 24, 78, 79, 86] bring out the strategy of discrete power assignment to have energy efficient routing topology. So when it comes to power assignment to nodes, work has been limited to discrete power management in topology control. Although some TC techniques [62] involve dynamic adjustment of transmission power level of nodes but in an implicit way to reduce power consumption and extend network lifetime. Authors in [81] come-up with interval graph based polynomial time algorithms to explicitly deal with nodes having power levels in the form of intervals. However this work is limited to static WSNs and algorithms are centralized in nature. Therefore more work on TC with continuous range assignment can be explored further. Also existing solutions to

achieve minimum cost routing topology by minimizing nodes' power consumption or engaging smaller transmission range enhance transmission delay in large network so work to obtain minimum delay energy efficient topology can be promoted in future work.

Interference minimizing models and approaches Approaches in [15, 17, 36, 37, 88, 89, 91, 93] are good enough to minimize nodes energy consumption and maximize network lifetime by explicitly reducing network interference but topology formed by them are less fault-tolerant and non-dynamic. Bao and Deng in 2016 [38] introduced the fault-tolerant topology with low interference. Most of the existing interference models work with static transmission range of nodes in minimizing interference and energy consumption [14, 15]. Appropriate estimation and reduction of interference in mobile WSNs is still an existing issue in this field.

8.2 Routing techniques

Classical hierarchical routing protocols These protocols or routing techniques have received maximum attention for reducing network energy consumption with incorporation of scalability factor. LEACH, Extension of LEACH and many more multi-hop cluster-based routing protocols [20] have been designed undertaking the issues of energy consumption, scalability, network lifetime, reliability and load balancing yet the energy-hole problem has been mitigated by proposing unequal clustering protocols namely DUCR [113], HEEMP [119], ISCTO [125], HDDS [133] and fault tolerant issue by CRPD [117], CFTR [96], EEMCS [129], ECRP [131], SHSDA [135]. Yet SHSDA [135] is the most secured, reliable and stable among the existing cluster-based routing protocol for static and homogenous WSNs. Further, HDDS [133] has better network lifetime and coverage than DUCR and CRPD. CFTR [96] conserves more energy, more stable and fault-tolerant than DUCR protocol. EEMCS [129] does $20-30\%$ better energy utilization than CRPD and outperforms LEACH and MOD-LEACH in load balancing, network stability, network lifetime and throughput in large-scale network. Also EESRA [123] performs better than LEACH and its variant in term of energy consumption, load balancing and network lifetime. The SEA [137] approach outperforms LEACH, SEP, TEEN and DEEC protocols in term of CHs selection, CHs rotation and network lifetime compared to . Tree-structure based data aggregation method given in [134] offers the best network lifetime and reliability with reduced energy consumption and average delay over the existing tree-based data aggregation and clustering approaches. Mostly available works have addressed energy optimization, coverage preservation and fault tolerance aspects independently. Also developed protocols are lacking in real-time application based WSNs because of persisting transmission delay while establishing network topology. For heterogenous large scale mobile network, more energy and fault-aware routings need to develop.

PSO-based algorithms Algorithm in [149] outperforms many Genetic Algorithm (GA) based on clustering schemes, namely GADA-LEACH, GABEEC, DCH-GA in term of minimizing energy consumption, maximizing throughput and scalability and has better applicability to real-time communication. Also inadequate clustering technique of PSOBS [147] has been overcome by Kaur et al. in [149]. Kaswan et al. [144] through their algorithm MOPSO overcome the issue of unbalanced energy consumption and load distribution of 'Tree cluster based data gathering algorithm (TCBDGA)' and 'energy aware sink relocation (EASR)'. PUDCRP [145] balances the energy consumption of sensor nodes more effectively, has good scalability and consumes less energy per round than the existing PSO-based protocols namely PSO-ECHS and PSO-SD. EC-PSO [146] minimizes energy consumption and maximizes network lifetime better than VD-PSO. MDCH-PSO [153] outperforms the standard LEACH-M, LEACH-C algorithm by improving the average residual energy by 13.9% and 27% respectively. Also average delay has been reduced by 29.6385% and 35.26% than LEACH-M and LEACH-C by MDCH-PSO algorithm respectively. GAPSO-H [155] outperforms some existing PSO-based algorithms (PSOECSM, PSO-UFC, PSOBS) and GA-based algorithms (GADA-LEACH, GABEEC, DCH-GA) in term of stability, network lifetime and throughput.

ACO-based algorithms ACO-based EMPAR and MPAR schemes by Jari and Avokh [166] have better performance than the ECDA algorithm in terms of energy consumption and network lifetime. EMPAR reduces energy consumption by 5.98% and 12.20% and increases the network lifetime by 12.26% and 30.38% compared to MPAR algorithm and energy-aware CS-based data aggregation algorithm respectively. Although, EMPAR performs better than MPAR. Improved ACO-based algorithm [163] is better to extend network lifetime and residual energy than the randomly distributed WSN strategy and ACO strategy. Further AOSTEB algorithm [164] reports the enhancement of network lifetime by more than 200% as compared to improved-ACO (IACO). Roy et al. in 2021 [167] put-forth an ACO-based technique which comes out to be the most efficient in terms of network lifetime, energy consumption, packet delivery ratio and end-to-end delay as compared with the existing ones concerned with the issue of non-uniform and uniform nodes densities in target area.

Hybrid PSO and ACO techniques A hybrid algorithm of ACO and PSO in [160] outperforms Genetic algorithm (GA) and Articifial Bee Colony (ABC) algorithms in saving nodes energy and improving throughput. More work using PSO and ACO techniques can be conducted to optimize energy

consumption and bring network stability with delay minimization in heterogenous large-scale dynamic network. It has been observed that hybrid algorithms of ACO and PSO are having better performance therefore, research on bringing hybrid algorithms with distinct multi-objectives can be more helpful in this field. Most of the PSO and ACO based routing techniques engage mobile sinks while the other nodes can be static or mobile, this makes the network more costlier so work on same objectives with different approach to obtain a low cost network should be under taken in future.

Real-time routing protocols WSNs face an inherent conflict between energy efficient communication and desired quality of service such as end-to-end communication delay, timely data delivery and/or deadline miss ratio. The well established MMSPEED protocol performs better than SPEED and SPEED is better than RAP in the domain of timely packet delivery and reliability but lacking energy efficient communication as protocols are not concerned with nodes residual energy and loss of energy while transmission. THVR [58] which focuses on minimizing delay factor with load balancing, outperforms SPEED protocol in lowering end-to-end deadline miss ratio and utilizing energy efficiency. EE-ARP protocols in [59] improves the deadline miss ratio of THVR and PATH protocol. ERTLD [60] which is the extension of RTLD protocol minimizes average end-end delay upto 65% and 31% compared to RTLD and MMSPEED respectively. Moreover ERTLD has better packet delivery rate as compared to RTLD and MMSPEED. DCEM algorithm of Huyunh et al. [174] overcomes the latency effect of cluster-based Gen-LEACH and multihop-HEED protocols. MDET [182] outperforms tree-structure based data flooding schemes in minimizing transmission delay with load balancing. DCEMRA [61] has better packet delivery ratio with the reduction in energy consumption and transmission delay compared to LEACH and PEGASIS. Real-time routing protocols meet the delay and QoS requirements by ignoring energy saving and lifetime prolonging factors. Therefore work on increasing network lifetime with minimum deadline miss ratio needs to undertake since node's battery gets drained so fast. Real-time routing protocols can utilize multiple paths to achieve energy efficiency and delay minimization simultaneously in future work. Also work on improving energy efficiency is achieved with using two-hop forwarding method only, it can be extended further.

9 Conclusion

In this paper, a detail review of 145 articles from the respected journals and conferences representing the topology control methods and routing techniques in WSNs have been provided. This review article offers the literature of graph-based methods utilized in range assignment problem, algorithmic approaches to minimize interference explicitly in topology control and various routing protocols/ techniques related to classical hierarchical routing, real-time routing, PSO and ACO-based routing with parallel mentioning of contribution made by the proposed solution, detail of methodology opted and limitations which might be overcome in future work. A brief review on previously existing survey papers has also been conducted with the observation that the existing review articles on topology control methods skipped many models and algorithms, particularly related to interference minimization in topology control. Also no survey paper has provided the combined and collective study of topology control methods and routing techniques. Therefore, reviewing unexplored and latest topology control methods and routing techniques in WSNs, dealing with various topology control and routing challenges, and placing them together systematically is the accomplished task of this review paper. Moreover, research methodology to conduct literature survey, from selection criteria to papers filtering, has been offered for readers reference. Henceforth, this review article acknowledges readers about the application of WSNs, existing issues related to topology control of WSNs, routing challenges, accomplished objectives of the existing solutions along with the detail of methodology opted, results, merits and demerits, a brief comparison among the proposed solutions, research gaps/ open issues and direction of future research work.

Declarations

Conflicts of interests The authors of the paper declare that they do not have any conflict of interest.

References

1. Matin MA, Islam MM (2012) Overview of wireless sensor network. Wireless Sensor Networks-Technology and Protocols 1-3
2. Al-Karaki JN, Kamal AE (2004) Routing techniques in wireless sensor networks: A survey. IEEE Wirel Commun 11:6–28
3. Akyildiz IF, Su W, Sankarasubramaniam Y, Cayirci E (2002) Wireless sensor networks: A survey. Comput Netw 38:393–422
4. Yick J, Mukherjee B, Ghosal D (2008) Wireless sensor network survey. Comput Netw 52:2292–2330
5. Al-Aghbari Z, Khedr AM, Osamy W, Arif I, Agrawal DP (2019) routing in wireless sensor networks using optimization techniques: A survey. Wirel Pers Commun 1-28
6. Ketshabetswe LK, Zungeru AM, Mangwala M, Chuma JM, Sigweni B (2019) Communication protocols for wireless sensor networks: A survey and comparison. Heliyon 5:e01591
7. Lai X, Ji X, Zhou X, Chen L (2017) Energy efficient link-delay aware routing in wireless sensor networks. IEEE Sens J 18:837–848
8. Maurya S, Jain VK, Chowdhury DR (2019) Delay aware energy efficient reliable routing for data transmission in heterogeneous mobile sink wireless sensor network. J Netw Comput Appl 144:118–137

9. Fanian F, Rafsanjani MK (2019) Cluster-based routing protocols in wireless sensor networks: A survey based on methodology. J Netw Comput Appl 142:111–142
10. Lin D, Wang Q et al (2020) A survey on energy-efficient strategies in static wireless sensor networks. ACM Transactions on Sensor Networks (TOSN) 17:1–48
11. Lu C, Blum BM et al (2002) Rap: A real-time communication architecture for large-scale wireless sensor networks. In: Eighth IEEE Real-Time and Embedded Technology and Applications Symposium, IEEE, pp 55-66
12. He T, Stankovic JA, Lu C, Abdelzaher T (2003) Speed: A stateless protocol for real-time communication in sensor networks. In: 23rd International Conference on Distributed Computing Systems, IEEE, pp 46-55
13. Santi P (2005) Topology control in wireless ad hoc and sensor networks. ACM computing surveys (CSUR) 37:164–194
14. Rickenbach PV, Wattenhofer R, Zollinger A (2009) Algorithmic models of interference in wireless ad hoc and sensor networks. IEEE/ACM Trans Networking 17:172–185
15. Panda BS, Shetty DP (2013) Minimum interference strong bidirectional topology for wireless sensor networks. Int J Ad Hoc Ubiquitous Comput 13:243–253
16. Shetty DP, Lakshmi MP (2019) Approximation algorithm for receiver interference problem in dual power wireless sensor networks. J Appl Math Comput 61:87–99
17. Lakshmi MP, Shetty DP et al (2019) Optimal algorithm for minimizing interference with two power levels in wireless sensor networks. J Commun 14:1198–1204
18. Singla P, Munjal A (2020) Topology control algorithms for wireless sensor networks: A review. Wireless Pers Commun 113:2363–2385
19. Kim B-S, Park HS, Kim KH, Godfrey D, Kim K-I (2017) A survey on real-time communications in wireless sensor networks. Wirel Commun Mob Comput 2017
20. Chan L, Chavez KG, Rudolph H, Hourani A (2020) Hierarchical routing protocols for wireless sensor network: a compressive survey. Wirel Netw 1-24
21. Maryem M, Belkassem T et al (2020) Routing in wireless sensor networks using fuzzy logic: A survey. In: 2020 International Conference on Intelligent Systems and Computer Vision (ISCV), IEEE, pp 1-6
22. Moaveninejad K, Li X-Y (2005) Low-interference topology control for wireless ad hoc networks. Ad Hoc Sens Wirel Networks 1:41–64
23. Yuanyuan Z, Jia X, Yanxiang H (2006) Energy efficient distributed connected dominating sets construction in wireless sensor networks. In: Proceedings of the 2006 international conference on Wireless communications and mobile computing, pp 797-802
24. Abu-Affash AK, Carmi P, Tzur AP (2018) Dual power assignment via second hamiltonian cycle. J Comput Syst Sci 93:41–53
25. Abderrahim M, Hakim H, Boujemaa H, Touati F (2019) Energy-efficient transmission technique based on dijkstra algorithm for decreasing energy consumption in WSNs. In: 2019 19th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA), IEEE, pp 599-604
26. Hai-Bo Y, Wen-Yu C (2008) Distributed power control algorithm with multi-qos constraints for wireless sensor networks. 2008 IEEE International Conference on Networking. Sensing and Control, IEEE, pp 1031–1036
27. Cheng X, Narahari B et al (2003) Strong minimum energy topology in wireless sensor networks: Np-completeness and heuristics. IEEE Trans Mob Comput 2:248–256
28. Li X-Y, Wan P-J, Wang Y (2001) Power efficient and sparse spanner for wireless ad hoc networks. In: Proceedings Tenth International Conference on Computer Communications and Networks, IEEE, pp 564-567
29. Wang K, Wang L, Cai S, Qu S (2009) An energy-saving algorithm of wsn based on gabriel graph. 2009 5th International Conference on Wireless Communications. Networking and Mobile Computing, IEEE, pp 1–4
30. Zhao H, Guo S, Wang X, Wang F (2015) Energy-efficient topology control algorithm for maximizing network lifetime in wireless sensor networks with mobile sink. Appl Soft Comput 34:539–550
31. Hong Z, Wang R, Wang N et al (2017) A tree-based topology construction algorithm with probability distribution and competition in the same layer for wireless sensor network. Peer-to-Peer Networking and Applications 10:658–669
32. Saranya V, Shankar S, Kanagachidambaresan GR (2018) Energy efficient clustering scheme (eecs) for wireless sensor network with mobile sink. Wireless Pers Commun 100:1553–1567
33. Chou C-H, Ssu K-F, Jiau HC et al (2010) A dead-end free topology maintenance protocol for geographic forwarding in wireless sensor networks. IEEE Trans Comput 60:1610–1621
34. Bagci H, Korpeoglu I, Yazici A (2014) A distributed fault-tolerant topology control algorithm for heterogeneous wireless sensor networks. IEEE Trans Parallel Distrib Syst 26:914–923
35. Zhang B, Jiao Z et al (2016) Efficient location-based topology control algorithms for wireless ad hoc and sensor networks. Wirel Commun Mob Comput 16:1943–1955
36. Sharma AK, Thakral N, Udgata SK, Pujari AK (2009) Heuristics for minimizing interference in sensor networks. In: International Conference on Distributed Computing and Networking, pp 49-54
37. Sun G, Zhao L, Chen Z, Qiao G (2015) Effective link interference model in topology control of wireless ad hoc and sensor networks. J Netw Comput Appl 52:69–78
38. Bao X, Deng C (2016) Fictc: Fault-tolerance-and-interference-aware topology control for wireless multi-hop networks. EURASIP J Wirel Commun Netw 2016:1–13
39. Kumar A, Pahuja S (2014) A comparative study of flooding protocol and gossiping protocol in wsn. Int J Comput Technol Appl 5:797–800
40. Heinzelman WR, Kulik J, Balakrishnan H (1999) Adaptive protocols for information dissemination in wireless sensor networks. In: Proceedings of the 5th annual ACM/IEEE International Conference On Mobile Computing and Networking, pp 174-185
41. Intanagonwiwat C, Govindan R et al (2003) Directed diffusion for wireless sensor networking. IEEE/ACM Trans Networking 11:2–16
42. Braginsky D, Estrin D (2002) Rumor routing algorthim for sensor networks. In: Proceedings of the 1st ACM international workshop on Wireless sensor networks and applications, pp 22-31
43. Schurgers C, Srivastava MB (2001) Energy efficient routing in wireless sensor networks. In: 2001 MILCOM Proceedings Communications for Network-Centric Operations: Creating the Information Force, IEEE, vol 1, pp 357-361
44. Yao Y, Gehrke J (2002) The cougar approach to in-network query processing in sensor networks. ACM SIGMOD Rec 31:9–18
45. Sadagopan N, Krishnamachari B, Helmy A (2003) The acquire mechanism for efficient querying in sensor networks. In: Proceedings of the First IEEE International Workshop on Sensor Network Protocols and Applications, IEEE, pp 149-155
46. Ye F, Chen A, Lu S, Zhang L (2001) A scalable solution to minimum cost forwarding in large sensor networks. In: Proceedings Tenth International Conference on Computer Communications and Networks, IEEE, pp 304-309
47. Yu Y, Govindan R, Estrin D (2001) Geographical and energy aware routing: A recursive data dissemination protocol for wireless sensor networks
48. Xu Y, Heidemann J, Estrin D (2001) Geography-informed energy conservation for ad hoc routing. In: Proceedings of the 7th annual international conference on Mobile computing and networking, pp 70-84
49. Li L, Halpern JY (2001) Minimum-energy mobile wireless networks revisited. In: ICC 2001 IEEE International Conference on Communications, IEEE, vol 1, pp 278-283

50. Elrahim AGA, Elsayed HA, Ramly SE, Ibrahim MM (2010) An energy aware wsn geographic routing protocol. Universal Journal of Computer Science and Engineering Technology 1:105–111
51. Sammut E, Debono CJ (2015) A location-based routing algorithm for wireless sensor networks. In: IEEE EUROCON 2015 International Conference on Computer as a Tool (EUROCON), IEEE, pp 1-5
52. Hu X, Ma L et al (2019) Fuzzy logic-based geographic routing protocol for dynamic wireless sensor networks. Sensors 19:196
53. Manuel AJ, Deverajan GG, Patan R, Gandomi AH (2020) Optimization of routing-based clustering approaches in wireless sensor network: Review and open research issues. Electronics 9:1630
54. Singh A, Sharma S, Singh J (2021) Nature-inspired algorithms for wireless sensor networks: A comprehensive survey. Computer Science Review 39:100342
55. Cao L, Cai Y, Yue Y (2019) Swarm intelligence-based performance optimization for mobile wireless sensor networks: Survey, challenges, and future directions. IEEE Access 7:161524–161553
56. Felemban E, Lee C-G, Ekici E (2006) Mmspeed: Multipath multispeed protocol for qos guarantee of reliability and timeliness in wireless sensor networks. IEEE Trans Mob Comput 5:738–754
57. Chipara O, He Z, Xing G et al (2006) Real-time power-aware routing in sensor networks. In: 2006 14th IEEE International Workshop on Quality of Service, IEEE, pp 83-92
58. Li Y, Chen CS, Song Y-Q et al (2009) Enhancing real-time delivery in wireless sensor networks with two-hop information. IEEE Trans Industr Inf 5:113–122
59. Rachamalla S, Kancherla AS (2016) A two-hop based adaptive routing protocol for real-time wireless sensor networks. Springerplus 5:1–12
60. Ahmed AA (2013) An enhanced real-time routing protocol with load distribution for mobile wireless sensor networks. Comput Netw 57:1459–1473
61. Selvi M, Velvizhy P, Ganapathy S et al (2019) A rule based delay constrained energy efficient routing technique for wireless sensor networks. Clust Comput 22:10839–10848
62. Aziz AA, Sekercioglu YA, Fitzpatrick P, Ivanovich M (2012) A survey on distributed topology control techniques for extending the lifetime of battery powered wireless sensor networks. IEEE Communications Surveys Tutorials 15:121–144
63. Yetgin H, Cheung KTK, El-Hajjar M, Hanzo LH (2017) A survey of network lifetime maximization techniques in wireless sensor networks. IEEE Communications Surveys Tutorials 19:828–854
64. Mohamed RE, Saleh AI, Abdelrazzak M, Samra AS (2018) Survey on wireless sensor network applications and energy efficient routing protocols. Wireless Pers Commun 101:1019–1055
65. Nguyen L, Nguyen HT (2020) Mobility based network lifetime in wireless sensor networks: A review. Comput Netw 174:107236
66. Bhattacharyya D, Kim T-H, Pal S (2010) A comparative study of wireless sensor networks and their routing protocols. Sensors 10:10506–10523
67. Pantazis NA, Nikolidakis SA, Vergados DD (2013) Energy-efficient routing protocols in wireless sensor networks: A survey. IEEE Communications surveys and tutorials 15:551–591
68. Sabor N, Sasaki S, Abo-Zahhad M (2017) Ahmed SM (2017) A comprehensive survey on hierarchical-based routing protocols for mobile wireless sensor networks: Review, taxonomy, and future directions. Wirel Commun Mob Comput
69. Ogundile OO, Alfa AS (2017) A survey on an energy-efficient and energy-balanced routing protocol for wireless sensor networks. Sensors 17:1084
70. Guleria K, Verma AK (2019) Comprehensive review for energy efficient hierarchical routing protocols on wireless sensor networks. Wireless Netw 25:1159–1183
71. Rawat P, Chauhan S (2021) Clustering protocols in wireless sensor network: A survey, classification, issues, and future directions. Computer Science Review 40:100396
72. Zagrouba R, Kardi A (2021) Comparative study of energy efficient routing techniques in wireless sensor networks. Information 12:42
73. Rachamalla S, Kancharla AS (2013) A survey of real-time routing protocols for wireless sensor networks. International Journal of Computer Science and Engineering Survey 4:35
74. Liu X (2017) Routing protocols based on ant colony optimization in wireless sensor networks: A survey. IEEE Access 5:26303–26317
75. Chen X, Yu L, Wang T et al (2020) Artificial intelligence-empowered path selection: A survey of ant colony optimization for static and mobile sensor networks. IEEE Access 8:71497–71511
76. Rong Y, Choi H, Choi H-A (2004) Dual power management for network connectivity in wireless sensor networks. In: 18th International Parallel and Distributed Processing Symposium, IEEE, pp 225
77. Liu Y, Ni L, Hu C (2012) A generalized probabilistic topology control for wireless sensor networks. IEEE J Sel Areas Commun 30:1780–1788
78. Lam NX, Nguyen TN, An MK, Huynh DT (2015) Dual power assignment optimization and fault tolerance in wsns. J Comb Optim 30:120–138
79. Sisodiya N, Shetty DP (2015) Total power minimization using dual power assignment in wireless sensor networks. In: 2015 International Conference on Information Technology (ICIT), IEEE, pp 26-30
80. Panda BS, Shetty DP, Pandey A (2015) K-distinct strong minimum energy topology problem in wireless sensor networks. In: International Conference on Distributed Computing and Internet Technology, Springer, pp 187-192
81. Kavra R, Gupta A, Kansal S (2021) Interval graph based energy efficient routing scheme for a connected topology in wireless sensor networks. Wireless Networks 1-20
82. Jain K (2001) A factor 2 approximation algorithm for the generalized steiner network problem. Combinatorica 21:39–60
83. Li X, Cai J, Zhang H (2016) Topology control for guaranteed connectivity provisioning in heterogeneous sensor networks. IEEE Sens J 16:5060–5071
84. Panda BS, Bhatta BK, Mishra D, De S (2017) New heuristics for strong minimum energy topology with reduced time complexity. In: 2017 IEEE International Conference on Advanced Networks and Telecommunications Systems (ANTS), pp 1-6
85. Mir ZH, Ko YB (2017) Collaborative topology control for many-to-one communications in wireless sensor networks. IEEE Access 5:15927–15941
86. Lakshmi MP et al (2019) Minimizing the total range with two power levels in wireless sensor networks. In Advanced Computing and Communication Technologies 183–191
87. Song L, Liu C et al (2019) Minimum connected dominating set under routing cost constraint in wireless sensor networks with different transmission ranges. IEEE/ACM Trans Networking 27:546–559
88. Wu Y, Hu Y et al (2018) Topology control for minimizing interference with delay constraints in an ad hoc network. J Parallel Distrib Comput 113:63–76
89. Lou T, Tan H, Wang Y, Lau FCM (2011) Minimizing average interference through topology control. International Symposium on Algorithms and Experiments for Sensor Systems. Springer, Wireless Networks and Distributed Robotics, pp 115–129
90. Tan H, Lou T, Wang Y et al (2011) Exact algorithms to minimize interference in wireless sensor networks. Theoret Comput Sci 412:6913–6925

91. Chiwewe TM, Hancke GP (2011) A distributed topology control technique for low interference and energy efficiency in wireless sensor networks. IEEE Trans Industr Inf 8:11–19
92. Korman M (2012) Minimizing interference in ad hoc networks with bounded communication radius. Inf Process Lett 112:748–752
93. Shetty DP, Lakshmi MP (2019) Minimizing the maximum sender interference by deploying additional nodes in a wireless sen-sor network. Electronic Journal of Graph Theory and Applications 7:169–182
94. Alsaafin A, Khedr AM, Aghbari ZA (2018) Distributed trajectory design for data gathering using mobile sink in wireless sensor networks. AEU-International Journal of Electronics and Communications 96:1–12
95. Gao Y, Wang J, Wu W et al (2019) Travel route planning with optimal coverage in difficult wireless sensor network environment. Sensors 19:1838
96. Moussa N, El Alaoui AEB (2019) A cluster-based fault-tolerant routing protocol for wireless sensor networks. Int J Commun Syst 32:e4131
97. Yang L, Lu YZ et al (2018) An unequal cluster-based routing scheme for multi-level heterogeneous wireless sensor networks. Telecommun Syst 68:11–26
98. Heinzelman WB, Chandrakasan AP, Balakrishnan H (2002) An application-specific protocol architecture for wireless microsensor networks. IEEE Trans Wireless Commun 1:660–670
99. Lindsey S, Raghavendra C, Sivalingam KM (2002) Data gathering algorithms in sensor networks using energy metrics. IEEE Trans Parallel Distrib Syst 13:924–935
100. Manjeshwar A, Agrawal DP (2001) Teen: Arouting protocol for enhanced efficiency in wireless sensor networks. In: ipdps 1:189
101. Baghyalakshmi D, Ebenezer J, Satyamurty SAV (2010) Low latency and energy efficient routing protocols for wireless sensor networks. In: 2010 International Conference on Wireless Communication and Sensor Computing (ICWCSC), IEEE, pp 1-6
102. Mo HS, Lee E, Park S, Kim S-H (2013) Virtual line-based data dissemination for mobile sink groups in wireless sensor networks. IEEE Commun Lett 17:1864–1867
103. Al-Karaki JN, Kamal AE (2008) Efficient virtual-backbone routing in mobile ad hoc networks. Comput Netw 52:327–350
104. Kusdaryono A, Lee K-H (2011) A clustering protocol with mode selection for wireless sensor network. Journal of Information Processing Systems 7:29–42
105. Xi-rong B, Shi Z, Ding-yu X, Zhi-tao Q (2010) An energy-balanced chain-cluster routing protocol for wireless sensor networks. In: 2010 Second International Conference on Networks Security, Wireless Communications and Trusted Computing, IEEE, vol 2, pp 79-84
106. Chen K-H, Huang J-M, Hsiao C-C (2009) Chiron: An energy-efficient chain-based hierarchical routing protocol in wireless sensor networks. In: 2009 Wireless Telecommunications Symposium, IEEE, pp 1-5
107. Sangolgi NB, Zakir SKA (2013) Energy aware data aggregation technique in wsn. Int J Sci Res Publ 376
108. Kim H-S, Han K-J (2005) A power efficient routing protocol based on balanced tree in wireless sensor networks. In: First International Conference on Distributed Frameworks for Multimedia Applications, IEEE, pp 138-143
109. Qiu W, Skafidas E, Hao P (2009) Enhanced tree routing for wireless sensor networks. Ad Hoc Netw 7:638–650
110. Singh S, Chand S, Kumar R et al (2016) Neecp: Novel energy-efficient clustering protocol for prolonging lifetime of wsns. IET Wireless Sensor Systems 6:151–157
111. Banimelhem O, Khasawneh S (2009) Grid-based multi-path with congestion avoidance routing (gmcar) protocol for wireless sensor networks. In: 2009 International Conference on Telecommunications, IEEE, pp 131-136
112. Tan ND, Viet ND (2015) Sstbc: Sleep scheduled and tree-based clustering routing protocol for energy-efficient in wireless sensor networks. In: The 2015 IEEE RIVF International Conference on Computing Communication Technologies-Research, Innovation, and Vision for Future (RIVF), IEEE, pp 180-185
113. Mazumdar N, Om H (2017) Ducr: Distributed unequal cluster-based routing algorithm for heterogeneous wireless sensor networks. Int J Commun Syst 30:e3374
114. Xin H, Liu X (2017) Energy-balanced transmission with accurate distances for strip-based wireless sensor networks. IEEE Access 5:16193–16204
115. Wen W, Zhao S, Shang C, Chang CY (2017) EAPC: Energy-aware path construction for data collection using mobile sink in wireless sensor networks. IEEE Sens J 2:890–901
116. Movva P, Rao PT (2018) Novel two-fold data aggregation and MAC scheduling to support energy efficient routing in wireless sensor network. IEEE Access 7:1260–1274
117. Wang S, Yu J et al (2018) Crpd: a novel clustering routing protocol for dynamic wireless sensor networks. Pers Ubiquit Comput 22:545–559
118. Dutt S, Agrawal S, Vig R (2018) Cluster-head restricted energy efficient protocol (CREEP) for routing in heterogeneous wireless sensor networks. Wireless Pers Commun 100:1477–1497
119. Sajwan M, Gosain D, Sharma AK (2018) Hybrid energy-efficient multi-path routing for wireless sensor networks. Comput Electr Eng 67:96–113
120. Yarinezhad R, Sarabi A (2018) Reducing delay and energy consumption in wireless sensor networks by making virtual grid infrastructure and using mobile sink. AEU-International Journal of Electronics and Communications 84:144–152
121. Agrawal A, Singh V, Jain S, Gupta RK (2018) GCRP: Grid-cycle routing protocol for wireless sensor network with mobile sink. AEU-International Journal of Electronics and Communications 94:1–11
122. Vahabi S, Eslaminejad M, Dashti SE (2019) Integration of geographic and hierarchical routing protocols for energy saving in wireless sensor networks with mobile sink. Wireless Netw 25:2953–2961
123. Elsmany EFA, Omar MA et al (2019) EESRA: Energy efficient scalable routing algorithm for wireless sensor networks. IEEE Access 7:96974–96983
124. Sha C, Song D et al (2019) A type of energy-balanced tree based data collection strategy for sensor network with mobile sink. IEEE Access 7:85226–85240
125. Park J, Kim S, Youn J et al (2020) Iterative sensor clustering and mobile sink trajectory optimization for wireless sensor network with nonuniform density. Wirel Commun Mob Comput
126. Wen W, Shang C et al (2020) DEDC: Joint density-aware and energy-limited path construction for data collection using mobile sink in WSNs. IEEE Access 8:78942–78955
127. Wang K, Yu CM, Wang LC (2020) DORA: A Destination-Oriented Routing Algorithm for Energy-Balanced Wireless Sensor Networks. IEEE Internet Things J 8:2080–2081
128. Mehta D, Saxena S (2020) MCH-EOR: Multi-objective cluster head based energy-aware optimized routing algorithm in wireless sensor networks. Sustainable Computing: Informatics and Systems 28:100406
129. Umbreen S, Shehzad D, Shafi N et al (2020) An energy-efficient mobility-based cluster head selection for lifetime enhancement of wireless sensor networks. IEEE Access 8:207779–207793
130. Mosavifard A, Barati H (2020) An energy-aware clustering and two-level routing method in wireless sensor networks. Computing 102:1653–1671
131. Moussa N, Hamidi-Alaoui Z, El Alaoui AEB (2020) Ecrp: An energy-aware cluster-based routing protocol for wireless sensor networks. Wirel Netw 1-14

132. Rawat P, Chauhan S, Priyadarshi R (2020) A novel heterogeneous clustering protocol for lifetime maximization of wireless sensor network. Wirel Pers Commun 1-17
133. Mazumdar N, Nag A, Nandi S (2021) Hdds: Hierarchical data dissemination strategy for energy optimization in dynamic wireless sensor network under harsh environments. Ad Hoc Netw 111:102348
134. Hasheminejad E, Barati H (2021) A reliable tree-based data aggregation method in wireless sensor networks. Peer-to-Peer Networking and Applications 1-15
135. Naghibi M, Barati H (2021) Shsda: secure hybrid structure data aggregation method in wireless sensor networks. J Ambient Intell Humaniz Comput 1-20
136. Yong J, Lin Z et al (2021) Tree-Based Multihop Routing Method for Energy Efficiency of Wireless Sensor Networks. Journal of Sensors
137. AL-Kaseem BR, Taha ZK et al (2021) Optimized Energy Efficient Path Planning Strategy in WSN with Multiple Mobile Sinks. IEEE Access
138. Liu J, Su S, Lu Y, Dong J (2021) A competition-based unequal clustering multihop approach for wireless sensor networks. Security and Communication Networks
139. Jain SK, Venkatadari M, Shrivastava N et al (2021) NHCDRA: a non-uniform hierarchical clustering with dynamic route adjustment for mobile sink based heterogeneous wireless sensor networks. Wireless Netw 27:2451–2467
140. Kumar S, Gautam PR, Rashid T et al (2021) Division algorithm based energy-efficient routing in wireless sensor networks. Wireless Pers Commun 1–20
141. Kaur T, Kumar D (2020) A survey on QoS mechanisms in WSN for computational intelligence based routing protocols. Wireless Netw 26:2465–2486
142. Azharuddin MD, Jana PK (2017) Pso-based approach for energy-efficient and energy-balanced routing and clustering in wireless sensor networks. Soft Comput 21:6825–6839
143. Kaur T, Kumar D (2018) Particle swarm optimization-based unequal and fault tolerant clustering protocol for wireless sensor networks. IEEE Sens J 18:4614–4622
144. Kaswan A, Singh V, Jana PK (2018) A multi-objective and pso based energy efficient path design for mobile sink in wireless sensor networks. Pervasive Mob Comput 46:122–136
145. Ruan D, Huang J (2019) A pso-based uneven dynamic clustering multi-hop routing protocol for wireless sensor networks. Sensors 19:1835
146. Wang J, Gao Y, Liu W et al (2019) An improved routing schema with special clustering using pso algorithm for heterogeneous wireless sensor network. Sensors 19:671
147. Tabibi S, Ghaffari A (2019) Energy-efficient routing mechanism for mobile sink in wireless sensor networks using particle swarm optimization algorithm. Wireless Pers Commun 104:199–216
148. Edla DR, Kongara MC, Cheruku R (2019) A pso based routing with novel fitness function for improving lifetime of wsns. Wireless Pers Commun 104:73–89
149. Kaur S, Grewal V (2020) A novel approach for particle swarm optimization-based clustering with dual sink mobility in wireless sensor network. Int J Commun Syst 33:e4553
150. Sahoo BM, Amgoth T, Pandey HM (2020) Particle swarm optimization based energy efficient clustering and sink mobility in heterogeneous wireless sensor network. Ad Hoc Netw 106:102237
151. Han Y, Byun H, Zhang L (2020) Energy-balanced cluster-routing protocol based on particle swarm optimization with five mutation operators for wireless sensor networks. Sensors 20:7217
152. Wang J, Gao Y, Zhou C et al (2020) Optimal coverage multi-path scheduling scheme with multiple mobile sinks for wsns. Computers, Materials Continua 62:695–711
153. Preethiya T, Muthukumar A, Durairaj S (2020) Double cluster head heterogeneous clustering for optimization in hybrid wireless sensor network. Wireless Pers Commun 110:1751–1768
154. Rawat P, Chauhan S (2020) Particle swarm optimization-based energy efficient clustering protocol in wireless sensor network. Neural Comput Applic 1-19
155. Sahoo BM, Pandey HM, Amgoth T (2021) Gapso-h: A hybrid approach towards optimizing the cluster based routing in wireless sensor network. Swarm Evol Comput 60:100772
156. Roy S, Mazumdar N, Pamula R (2021) An energy optimized and QoS concerned data gathering protocol for wireless sensor network using variable dimensional PSO. Ad Hoc Netw 123:102669
157. Sun Y, Dong W, Chen Y (2017) An improved routing algorithm based on ant colony optimization in wireless sensor networks. IEEE Commun Lett 21:1317–1320
158. Zhao Z, Hou M, Zhang N, Gao M (2017) Multipath routing algorithm based on ant colony optimization and energy awareness. Wireless Pers Commun 94:2937–2948
159. Kumar P, Amgoth T, Annavarapu CSR (2018) Aco-based mobile sink path determination for wireless sensor networks under non-uniform data constraints. Appl Soft Comput 69:528–540
160. Kaur S, Mahajan R (2018) Hybrid meta-heuristic optimization based energy efficient protocol for wireless sensor net- works. Egyptian Informatics Journal 19:145–150
161. Rathee M, Kumar S et al (2019) Ant colony optimization based quality of service aware energy balancing secure routing algorithm for wireless sensor networks. IEEE Trans Eng Manage 68:170–182
162. Li X, Keegan B, Mtenzi F et al (2019) Energy-efficient load balancing ant based routing algorithm for wireless sensor networks. IEEE Access 7:113182–113196
163. Chu K-C, Horng D-J, Chang K-C (2019) Numerical optimization of the energy consumption for wireless sensor networks based on an improved ant colony algorithm. IEEE Access 7:105562–105571
164. Arora VK, Sharma V, Sachdeva M (2019) Aco optimized self-organized tree-based energy balance algorithm for wireless sensor network. J Ambient Intell Humaniz Comput 10:4963–4975
165. Raj PP, Khedr AM, Aghbari ZA (2020) Data gathering via mobile sink in wsns using game theory and enhanced ant colony optimization. Wireless Netw 26:2983–2998
166. Jari A, Avokh A (2021) Pso-based sink placement and load-balanced anycast routing in multi-sink wsns considering compressive sensing theory. Eng Appl Artif Intell 100:104164
167. Roy S, Mazumdar N, Pamula R (2021) An optimal mobile sink sojourn location discovery approach for the energy-constrained and delay-sensitive wireless sensor network. J Ambient Intell Humaniz Comput 1-28
168. Moussa N, El Alaoui AEB (2021) An energy-efficient cluster-based routing protocol using unequal clustering and improved aco techniques for wsns. Peer-to-Peer Networking and Applications 1-14
169. Donta PK, Amgoth T, Annavarapu CSR (2021) An extended ACO-based mobile sink path determination in wireless sensor networks. J Ambient Intell Humaniz Comput 12:8991–9006
170. Mahapatra A, Anand K, Agrawal DP (2006) Qos and energy aware routing for real-time traffic in wireless sensor networks. Comput Commun 29:437–445
171. Ergen SC, Varaiya P (2007) Energy efficient routing with delay guarantee for sensor networks. Wireless Netw 13:679–690
172. Huynh T-T, Tran T-N, Tran C-H, Dinh-Duc A-V (2017) Delay constraint energy-efficient routing based on lagrange relaxation in wireless sensor networks. IET Wireless Sensor Systems 7:138–145
173. Ahmed AA, Fisal N (2008) A real-time routing protocol with load distribution in wireless sensor networks. Comput Commun 31:3190–3203

174. Huynh T-T, Dinh-Duc A-V, Tran C-H (2016) Delay-constrained energy-efficient cluster-based multi-hop routing in wireless sensor networks. Journal of Communications and Networks 18:580–588
175. Jain S, Pattanaik KK, Verma RK et al (2020) Delay-Aware Green Routing for Mobile-Sink-Based Wireless Sensor Networks. IEEE Internet Things J 8:4882–4892
176. Dutt S, Agrawal S, Vig R (2021) Delay-sensitive, reliable, energy-efficient, adaptive and mobility-aware (dream) routing protocol for wsns. Wirel Pers Commun 1-29
177. Liu X, Qiu T, Zhou X et al (2019) Latency-aware path planning for disconnected sensor networks with mobile sinks. IEEE Trans Industr Inf 16:350–361
178. Lai S, Ravindran B (2012) Least-latency routing over time-dependent wireless sensor networks. IEEE Trans Comput 62:969–983
179. Li Y, Chen H, Mo S, Liu H (2014) Optimal query-driven data forwarding for delay-sensitive wireless sensor networks. Wireless Pers Commun 77:41–62
180. Tita ED, Nwadiugwu WP, Lee JM, Kim DS (2021) Real-time optimizations in energy profiles and end-to-end delay in WSN using two-hop information. Comput Commun 172:169–182
181. Wu J, Chen Z, Wu J et al (2020) An energy efficient data transmission approach for low-duty-cycle wireless sensor networks. Peer-to-Peer Networking and Applications 13:255–268
182. Cheng L, Niu J, Luo C et al (2018) Towards minimum-delay and energy-efficient flooding in low-duty-cycle wireless sensor networks. Comput Netw 134:66–77
183. Akkaya K, Younis M (2004) Energy-aware delay-constrained routing in wireless sensor networks. Int J Commun Syst 17:663–687
184. Boughanmi N, Song YQ (2008) A new routing metric for satisfying both energy and delay constraints in wireless sensor networks. J Signal Process Syst 51:137–143

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



Ms. Radhika Kavra is currently pursuing Ph.D. from Department of Applied Mathematics, Delhi Technological University, New Delhi, India. She has done M.Sc. Mathematics from IIT Roorkee. Her research area is Graph and Optimization.



Dr. Anjana Gupta is a Professor in the Department of Applied Mathematics, Delhi Technological University, New Delhi, India. She received her Ph.D. degree from Mathematics Department, Delhi University. Her field of specialization includes Optimization Techniques, Fuzzy Logics, Multicriteria Group Decision Making Problems. She has more than 22 years of teaching experience and more than 40 research publications in well-reputed International and National Journals and many in International conferences.



Dr. Sangita Kansal Sangita Kansal is a Professor in the Department of Applied Mathematics, Delhi Technological University, New Delhi, India. She received her Ph.D. degree from IIT Delhi. Her field of specialization includes Graph theory and Petri nets. She has more than 26 years of teaching experience and more than 30 years of research experience. She has research publications in well-reputed International and National Journals.

TLMOTE: A Topic-based Language Modelling Approach for Text Oversampling

Arjun Choudhry, Seba Susan, Anmol Bansal, Anubhav Sharma
Delhi Technological University, New Delhi, India
{arjunchoudhry_2k18it031, sebasusan, anmolbansal_2k18it025, anubhavsharma_2k18it029}@dtu.ac.in

Abstract

Training machine learning and deep learning models on unbalanced datasets can lead to a bias portrayed by the models towards the majority classes. To tackle the problem of bias towards majority classes, researchers have presented various techniques to oversample the minority class data points. Most of the available state-of-the-art oversampling techniques generate artificial data points which cannot be comprehensibly understood by the reader, despite the synthetic data points generated being similar to the original minority class data points. In this work, we present Topic-based Language Modelling Approach for Text Oversampling (TLMOTE), a novel text oversampling technique for supervised learning from unbalanced datasets. TLMOTE improves upon previous approaches by generating data points which can be intelligibly understood by the reader, can relate to the main topics of the minority class, and introduces more variations to the synthetic data generated. We evaluate the efficacy of our approach on various tasks like Suggestion Mining SemEval 2019 Task 9 Subtasks A and B, SMS Spam Detection, and Sentiment Analysis. Experimental results verify that oversampling unbalanced datasets using TLMOTE yields a higher macro F1 score than with other oversampling techniques.

1 Introduction

Class imbalance is one of the most prominent issues surrounding supervised learning models for years, despite the significant improvements in classifier architectures and preprocessing approaches. Classifiers trained on unbalanced datasets tend to be biased towards classes with more data points, or the majority classes, reducing the efficacy of the classifier in predicting the correct label for a given input. To this end, significant work has been done to ameliorate the problem of class imbalance in supervised learning tasks. Chawla et al. (2002) presented *Synthetic Minority Oversampling Technique* (SMOTE), a minority class oversampling technique which works in the Euclidean space and has been widely applied to a wide variety of tasks.

Over the years, many variations of SMOTE have been introduced to improve the viability of data points generated for the minority classes. He et al. (2008) introduced *Adaptive Synthetic Sampling Approach for Imbalanced Learning* (ADASYN), a synthetic sampling approach which uses a weighted distribution for the minority classes based on the difficulty level in their learning. Unlike SMOTE, in which each point is given a uniform weight, in ADASYN, a density distribution decides the number of synthetic samples that need to be generated for a specific point. Han, Wang, and Mao (2005) proposed *Borderline-SMOTE*, a variation of SMOTE in which minority class data points whose neighbours are all majority class data points are considered as noise points, and are ignored during oversampling.

Some recent works have also used undersampling of the majority class, or hybrid of oversampling and undersampling approaches for tackling the problem of class imbalance. Susan and Kumar (2018) proposed a hybrid model of minority oversampling and Particle Swarm Optimization-based majority undersampling. These SMOTE-based approaches are applicable across multiple modalities. There have been some approaches proposed to tackle the problem of class imbalance for specific modalities as well. Saini and Susan (2019) proposed a minority oversampling approach for the classification of breast cancer histopathological images, by the generation of synthetic image samples using affine transformations. Moreo, Esuli, and Sebastiani (2016) proposed *Distributed Random Oversampling* (DRO), a text oversampling approach based on assigning a probabilistic generative function to each sample in the minority class of the training set. Anand et al. (2018) tackled class imbalance in Phishing URL Detection datasets using *Text Generative Adversarial Networks*, and generated synthetic URLs similar to those in the original dataset. Leekha, Goswami, and Jain (2020) proposed *Language Model-based Oversampling Technique* (LMOTE), a minority class oversampling technique exclusively for textual data. They found that LMOTE performed comparably to SMOTE, while allowing for a more intuitive understanding of the generated samples.

In this work, we introduce *Topic-based Language Modelling approach for Text Oversampling* (TLMOTE), an oversampling technique using Latent Dirichlet Allocation-based Topic Modelling and Language Modelling for the generation of synthetic data points for the minority classes. We evaluate TLMOTE against SMOTE and LMOTE on Suggestion Mining *SemEval* 2019 Task 9 Subtasks A and B, SMS Spam

Copyright © 2021by the authors. All rights reserved.

Detection, and Sentiment Analysis for a variety of machine learning and deep learning classifiers. The organization of this paper is as follows. TLMOTE is presented in Section 2, the experimental results are discussed in Section 3, and the conclusions are summarized in Section 4.

2 Methodology

2.1 Datasets and Preprocessing

We use the *SemEval* 2019 Task 9 Subtasks A and B datasets for Suggestion Mining (Negi, Daudert, and Buitelaar, 2019), SMS Spam Classification dataset (Almeida, Hidalgo, and Yamakami, 2011) made available in UCI machine learning repository, and the COVID-19 Vaccine Tweets[1] dataset for sentiment analysis. The Suggestion Mining Subtask A dataset consists of training and testing data from suggestion forums for Windows platform developers, while the Subtask B dataset consists of the same training data as Task A, and testing data from the hotel reviews domain. The Covid Vaccine Tweets dataset consists of tweets on various COVID-19 vaccines, annotated with their correct sentiments. The distributions of these datasets are shown in Table 1.

Class	Size
Sugg	2085
Non-Sugg	6415

Class	Size
Spam	747
Ham	4827

Class	Size
Neg	420
Neut	3680
Pos	1900

Table 1: Class-wise distribution for Suggestion Mining training dataset, SMS Spam Classification dataset, and COVID-19 Vaccine Tweets dataset for Sentiment Analysis.

The training set for both Suggestion Mining Subtasks A and B are same. During pre-processing, we convert the text to lower case, and de-contract verbs forms (eg. "I'll" to "I will"). We further remove all punctuation marks before training and testing the classifiers on the unbalanced and oversampled data.

2.2 Observed Shortcomings of LMOTE

Upon generating synthetic data points using LMOTE for the suggestions class in the suggestion mining train set, we observe that some of the instances generated are highly duplicative. This is especially true for synthetic data points generated using 5-grams with relatively lower frequency.

During the extraction of the most common 5-grams in the suggestions class, the frequency of these 5-grams sharply drops to 1, and we observe 5-grams that occur consecutively within a data point in the initial data. Oversampling using these 5-grams leads to the generation of multiple near-identical data points, which can lead to model over-fitting.

Some examples of successive n-grams generated by the n-gram generator which led to duplicate data points sampled by the language model, along with the generated synthetic samples, are shown in Table 2. We observed many cases of such instances across all datasets oversampled using LMOTE.

[1]https://www.kaggle.com/datasciencetool/covid19-vaccine-tweets-with-sentiment-annotation

5-gram	Synthetic datapoint generated by LMOTE
i would expect the store	i would expect the store to be capable of searching the partial words also or may be more accessible and the rating can be the only send from the camera using more either h264 or mpeg4
would expect the store to	would expect the store to be capable of searching the partial words also or may be more accessible and the rating can be the only send from the camera using more either h264 or mpeg4
expect the store to be	expect the store to be capable of searching the partial words also or may be more accessible and the rating can be the only send from the camera using more either h264 or mpeg4
the store to be capable	the store to be capable of searching the partial words also or may be more accessible and the rating can be the only send from the camera using more either h264 or mpeg4

Table 2: Examples of duplicity in the synthetic data points generated for the suggestion mining dataset using LMOTE, due to consecutively occurring 5-gram seed values.

2.3 TLMOTE Proposed Procedure

To reduce the duplicity of generated data points for the minority class, we generate 5-grams which consist of words that relate most to the prominent topics of the minority classes. These words, and thereby the 5-grams are extracted using topic modelling with Latent Dirichlet Allocation (Blei, Ng, and Jordan, 2003).

We further use three language models, which take the last 4 tokens, 5 tokens and 6 tokens, respectively, as the input to predict the next word. We randomly select one word out of the three predictions made by the language models as the next word in the sequence, and append it to the end of the sequence. This process continues till a random number of words between 20 and 30 are appended after the 5-gram. Using three language models randomly, instead of one, introduces more variation into the data points generated, even when similar 5-grams are taken as the seed.

For each minority class of a dataset, TLMOTE follows the following procedure to generate synthetic samples:

Find top α topics in the minority class We employ topic modelling using Latent Dirichlet Allocation for extracting the ($\alpha = 100$) most important topics in the minority class samples, using the Gensim package.

Extract the top 10 keywords K_t for each topic t We further extract the 10 most prominent keywords K_t for each topic t extracted, and create a list L consisting of the set of all top 10 keywords for each topic.

$$L = K_1 \cup K_2 \cup ...K_{100}$$

Extract the top η 5-grams from the minority class samples We provide L to the n-gram extractor, and extract η 5-grams, which contain one or more of the keywords in L. η equals the deficit of samples in the minority class as compared to the majority class. This ensures that sequences of 5-grams that occur consecutively do not get extracted in significant numbers, as the chain gets broken when a certain 5-gram does not contain a keyword from the list L.

Train three language models on the minority class samples We train three BiLSTM-based language models on

the minority class samples, which predict the probability distribution of the next word w_x over the entire vocabulary V of the minority class samples, based on the previous 4 words (w_{x-4}, w_{x-3}, w_{x-2}, w_{x-1}), 5 words (w_{x-5}, w_{x-4}, w_{x-3}, w_{x-2}, w_{x-1}) and 6 words (w_{x-6}, w_{x-5}, w_{x-4}, w_{x-3}, w_{x-2}, w_{x-1}), respectively. The language models are trained using GloVe (Pennington, Socher, and Manning, 2014) 100 dimension embeddings. The models predict the probability distributions $P(w_i \mid w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})\ \forall i \in (V)$, $P(w_i \mid w_{x-5}\ w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})\ \forall i \in (V)$, and $P(w_i \mid w_{x-6}\ w_{x-5}\ w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})\ \forall i \in (V)$, such that:

$$w_{x4} = \arg\max_{w_i} P(w_i \mid w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})$$

$$w_{x5} = \arg\max_{w_i} P(w_i \mid w_{x-5}\ w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})$$

$$w_{x6} = \arg\max_{w_i} P(w_i \mid w_{x-6}\ w_{x-5}\ w_{x-4}\ w_{x-3}\ w_{x-2}\ w_{x-1})$$

$$w_x = \textbf{random}(w_{x4}, w_{x5}, w_{x6})$$

Generate synthetic samples using the extracted 5-grams and the three language models We generate synthetic data points using the three language models and a 5-gram from the set η, by appending the next word w_x for a random number of times between 20 and 30. This process is repeated for all 5-grams in η.

The proposed algorithm for TLMOTE is as follows:

Algorithm 1 Topic-based Language Modelling Approach for Text Oversampling (TLMOTE)

Input:
$\mathcal{M}_s = \{r_i \in \mathcal{M} \mid n_i = \{`minority\ class\ sample'\}\}$ Samples from a minority class $\mathcal{M}$;
t- Number of topics to extract;
k- Number of most prominent keywords per topic;
η- Number of n-grams to use in TLMOTE;
n- type of n-grams *e.g.* 5 for 5-grams, etc.;
$\mathcal{N}$- Number of synthetic samples required.
Output: $\mathcal{S} : \mathcal{N}$ synthetic samples for the minority class

$keywords \leftarrow$ **TopicModel**($\mathcal{M}_s$, t, k)
$topic_n_grams \leftarrow$ **NGrams**($\mathcal{M}_s$, $keywords$, $n = 5$, $\mathcal{N}$)
$4gram_LM \leftarrow$ **TrainLanguageModel**($\mathcal{M}_s$, $n = 4$)
$5gram_LM \leftarrow$ **TrainLanguageModel**($\mathcal{M}_s$, $n = 5$)
$6gram_LM \leftarrow$ **TrainLanguageModel**($\mathcal{M}_s$, $n = 6$)
Initialize $\mathcal{S} \leftarrow \mathcal{M}_s$
while $|\mathcal{S}| < \mathcal{N}$ **do**
  $seed \leftarrow$ **SeedGenerate**($topic_n_grams$)
  $sample \leftarrow$ **TLMOTEGenerate**($seed$, $4gram_LM$, $5gram_LM$, $6gram_LM$)
  $\mathcal{S} \leftarrow \mathcal{S} \cup sample$
end while
return $\mathcal{S}$

Algorithm 1 illustrates our proposed TLMOTE oversampling approach. Some of the procedures used in the algorithm are described below:

TopicModel($\mathcal{M}_s$, t, k): It extracts the top t topics from the minority class samples, and generates the top k keywords by weightage for each topic. It returns the combined list of the most prominent keywords for all t topics.

NGrams($\mathcal{M}_s$, $keywords$, $n = 5$, $\eta = \mathcal{N}$): It returns the top $\mathcal{N}$ 5-grams from the minority class samples, which contain one or more *keywords* extracted by **TopicModel**.

TrainLanguageModel($\mathcal{M}_s$, n): It trains an n-gram BiLSTM-based language model on $\mathcal{M}_s$.

SeedGenerate($topic_based_n_grams$): It iterates over the list of *n_grams* provided by **NGrams** and returns the next one as seed for each synthetic sample to be generated.

TLMOTEGenerate($seed$, $4gram_LM$, $5gram_LM$, $6gram_LM$): It takes the 4-gram, 5-gram and 6-gram language models, and an n-gram *seed* provided by **SeedGenerate** as input, and generates a synthetic sample for a random length between 25 and 35 tokens as output. Algorithm 2 summarizes the randomised approach used using three language models for the generation of synthetic samples.

We have made the code for TLMOTE available online[2].

Algorithm 2 TLMOTEGenerate procedure from TLMOTE algorithm

Input:
$seed$ - n-gram to be used for generating synthetic sample;
$4gram_LM$- 4-gram Language Model;
$5gram_LM$- 5-gram Language Model;
$6gram_LM$- 6-gram Language Model;
t- Number of tokens used as input to the language models;
min- Minimum length of tokens to be appended to seed in synthetic sample;
max- Maximum length of tokens to be appended to seed in synthetic sample;
Output: $Sent : Synthetic\ sample\ generated$

$length \leftarrow$ **random**(min, max)
Initialize *Sent* $\leftarrow$ *seed*
while |*Sent*| < *length* **do**
  Initialize *words* $\leftarrow$ [ ]
  words $\leftarrow$ *words* $\cup$ $4gram_LM$(**Last_t_tokens**($Sent$, $t = 4$))
  words $\leftarrow$ *words* $\cup$ $5gram_LM$(**Last_t_tokens**($Sent$, $t = 5$))
  if |*Sent*| >= 6 **then**
    words $\leftarrow$ *words* $\cup$ $6gram_LM$(**Last_t_tokens**($Sent$, $t = 6$))
  end if
  $next_token \leftarrow$ **RandomWord**($words$)
  $Sent \leftarrow$ **Concatenate**($Sent$, $next_token$)
end while
return $Sent$

Some of the procedures used in Algorithm 2 are described below:

[2]https://github.com/Arjun7m/TLMOTE

random($min,\ max$): It generates a number between *min* and *max*.

Last_t_tokens($Sent,\ t$): It returns the last t tokens in the list *Sent*.

RandomWord($words$): It returns one of the tokens in the list $words$ randomly.

Concatenate($Sent,\ next_token$): It appends the generated *next_token* to the end of *Sent*.

2.4 Classification Models

We evaluate the efficacy of TLMOTE by training and evaluating a variety of machine learning and deep learning classifiers on the unbalanced datasets, as well as datasets oversampled using SMOTE, LMOTE, TLMOTE and one of its variations. Among machine learning classifiers, we train and evaluate Random Forest (RF), Support Vector Machine (SVM) and Logistic Regression (LR) classifiers, while among deep learning architectures, we train and evaluate CNN-LSTM (Zhou et al., 2015) neural network, Attention-based Convolutional Bi-LSTM neural network (ACBiLSTM) (Kamyab, Liu, and Adjeisah, 2021) and Attention-based Recurrent Convolutional neural network (ARC) (Guo et al., 2019).

We optimise all machine learning models using GridSearch, while all deep learning models are optimised using Early Stopping to prevent over-fitting.

3 Results and Discussion

The experiments were performed using Python 3.7 version software on an NVIDIA Tesla P100 GPU. We evaluated the performance of a variety of deep learning and machine learning classifiers trained on the unbalanced datasets, as well as datasets oversampled using SMOTE, LMOTE and TLMOTE (referred to as TLMOTE (R) in the tables). To assess the impact of using three language models in a randomised order, we also oversample the datasets using a variation of TLMOTE (referred to as TLMOTE (M) in the tables), which selects the token with the highest probability among the three outputs of the language models, instead of choosing randomly.

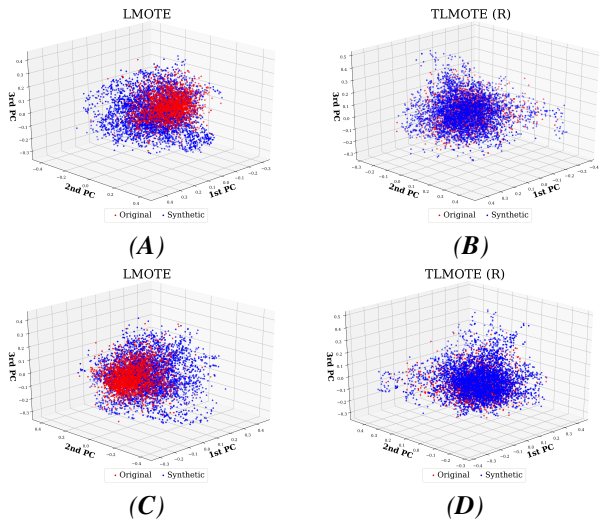


Figure 1: 3-dimensional PCA plots for the Suggestion Mining tasks oversampled using: (A) LMOTE (B) TLMOTE (C) LMOTE Reverse View (D) TLMOTE Reverse View.

Model	Class	Unbalanced	SMOTE	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1	F1
SVM	Non-suggestion	0.91	0.88	0.90	0.91	**0.92**
	Suggestion	0.33	0.38	0.36	**0.38**	**0.38**
	Macro avg	0.62	0.63	0.63	0.64	**0.65**
RF	Non-suggestion	**0.95**	0.91	0.94	0.94	0.94
	Suggestion	0.32	0.46	0.48	**0.51**	0.50
	Macro avg	0.64	0.69	0.71	**0.72**	**0.72**
LR	Non-suggestion	0.88	0.87	0.88	0.89	**0.90**
	Suggestion	0.29	0.31	0.33	0.36	**0.38**
	Macro avg	0.58	0.59	0.61	0.63	**0.64**

Table 3: Performance evaluation of various Machine Learning models trained on the SemEval 2019 Task 9 Subtask A.

Model	Class	Unbalanced	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1
CNN-LSTM	Non-suggestion	0.92	0.94	0.93	**0.95**
	Suggestion	0.49	0.51	0.55	**0.58**
	Macro avg	0.70	0.72	0.74	**0.76**
ACBiLSTM	Non-suggestion	0.93	**0.96**	0.94	0.95
	Suggestion	0.56	0.58	0.61	**0.62**
	Macro avg	0.75	0.77	0.78	**0.79**
ARC	Non-suggestion	**0.95**	0.94	**0.95**	**0.95**
	Suggestion	0.54	0.55	0.56	**0.58**
	Macro avg	0.74	0.75	0.75	**0.77**

Table 4: Performance evaluation of various Deep Learning models trained on the SemEval 2019 Task 9 Subtask A.

Model	Class	Unbalanced	SMOTE	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1	F1
SVM	Non-suggestion	**0.73**	**0.73**	0.72	0.72	0.71
	Suggestion	0.06	**0.21**	0.08	0.12	0.17
	Macro avg	0.39	**0.47**	0.40	0.42	0.44
RF	Non-suggestion	**0.74**	**0.74**	**0.74**	**0.74**	0.73
	Suggestion	0.04	0.10	0.10	0.10	**0.12**
	Macro avg	0.39	0.42	0.42	0.42	**0.43**
LR	Non-suggestion	0.71	0.71	0.71	**0.72**	0.71
	Suggestion	0.22	**0.25**	0.23	0.23	**0.25**
	Macro avg	0.46	**0.48**	0.47	**0.48**	**0.48**

Table 5: Performance evaluation of various Machine Learning models trained on the SemEval 2019 Task 9 Subtask B.

We use the provided train-test split for the Suggestion Mining tasks, and randomly split the other datasets in the ratio 80:20 for training and testing sets of the other datasets. For the sake of a fair comparison, we use the same training splits for all oversampling techniques. Performance evaluations for all combinations of classifiers and oversampling techniques using F1 score for each class, as well as Macro F1 score for the total testing set, are illustrated in Tables 3 and 4 for Suggestion Mining Subtask A, Tables 5 and 7 for Suggestion Mining Subtask B, Tables 8 and 9 for the SMS Spam Detection task, and Tables 10 and 11 for the Sentiment Analysis task. Figure 1 represents Principal Component Analysis (PCA) (Jolliffe, 2002) graphs for the oversampled training sets generated for the suggestion mining task using TLMOTE (R) and LMOTE, plotted using text embeddings generated using a Tf-IDf vectorizer fit on the original data points. Some prominent findings we observed are:

Classifiers trained on datasets oversampled using TLMOTE outperformed those trained on datasets oversam-

5-gram	Synthetic datapoint generated by TLMOTE
please make support for setting	please make support for setting set automationid on corewindows and messagedialog could be supported in windows 10 too popular apps are displayed to every phone
make support for setting set	make support for setting set automationid on corewindows and messagedialog should be added to uwp for wp images should be tailored to be in a specific way
support for setting set automationid	support for setting set automationid on corewindows and messagedialog should be supported in store apps like we have download media like for windows 78 for windows 8 plz see this

Table 6: Examples of synthetic data points generated using TLMOTE for the suggestion mining training set. Data points show variations, even for consecutively occurring 5-gram seed values.

Model	Class	Unbalanced	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1
CNN-LSTM	Non-suggestion	0.74	**0.75**	**0.75**	0.74
	Suggestion	0.25	0.27	0.33	**0.34**
	Macro avg	0.49	0.51	**0.54**	**0.54**
ACBiLSTM	Non-suggestion	**0.75**	**0.75**	**0.75**	**0.75**
	Suggestion	0.29	0.24	**0.32**	**0.32**
	Macro avg	0.52	0.50	**0.54**	0.53
ARC	Non-suggestion	**0.75**	0.74	**0.75**	**0.75**
	Suggestion	0.26	0.20	0.26	**0.28**
	Macro avg	0.50	0.47	0.50	**0.52**

Table 7: Performance evaluation of various Deep Learning models trained on the SemEval 2019 Task 9 Subtask B.

Model	Class	Unbalanced	SMOTE	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1	F1
SVM	Ham	0.98	**0.99**	**0.99**	**0.99**	**0.99**
	Spam	0.88	0.96	0.94	0.95	**0.97**
	Macro avg	0.93	**0.98**	0.96	0.97	**0.98**
RF	Ham	**0.99**	**0.99**	**0.99**	**0.99**	**0.99**
	Spam	0.90	0.92	0.94	**0.96**	**0.96**
	Macro avg	0.94	0.95	0.96	**0.98**	**0.98**
LR	Ham	0.98	**0.99**	**0.99**	**0.99**	**0.99**
	Spam	0.88	0.95	0.95	0.96	**0.97**
	Macro avg	0.93	0.97	0.97	**0.98**	**0.98**

Table 8: Performance evaluation of various Machine Learning models trained on the SMS Spam Filtering dataset.

pled using SMOTE and LMOTE We observed that classifiers trained using TLMOTE outperformed those trained using SMOTE and LMOTE in terms of the macro F1 score across nearly all datasets and model architectures. In most cases, we observed that classifiers trained on datasets oversampled using TLMOTE portrayed a 1-5 percent improvement in macro F1 score over their counterparts trained on datasets oversampled using SMOTE and LMOTE, indicating significantly reduced bias shown by the classifiers trained on datasets oversampled using TLMOTE towards the majority classes. We saw slightly greater improvements with TLMOTE, as compared to LMOTE, in case of deep learning models (Tables 4, 7, 9 and 11) when compared with machine learning models (Tables 3, 5, 8 and 10) on the same datasets. This can be attributed to over-fitting in case of deep learning models when trained on datasets oversampled using LMOTE. Due to increased variations in the instances generated by TLMOTE, deep learning models trained on datasets oversampled using TLMOTE avoid the issue of overfitting.

Oversampling with three language models in a randomized order leads to marginal improvements in the performance of the classifiers We observed that classifiers trained using TLMOTE with three language models, with 4, 5 and 6 tokens as input length, respectively, and used in a randomized order (referred to as TLMOTE (R) in the tables), outperform their counterparts trained on datasets oversampled using TLMOTE (M), where the next word is taken from that language model which gives the highest probability for a word in the probability distribution.

Model	Class	Unbalanced	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1
CNN-LSTM	Ham	0.99	**1.00**	**1.00**	**1.00**
	Spam	0.95	0.97	**0.98**	**0.98**
	Macro avg	0.97	0.98	**0.99**	**0.99**
ACBiLSTM	Ham	0.99	**1.00**	0.99	**1.00**
	Spam	0.96	**0.98**	0.97	**0.98**
	Macro avg	0.98	**0.99**	0.98	**0.99**
ARC	Ham	0.99	0.99	0.99	**1.00**
	Spam	0.95	**0.97**	**0.97**	**0.97**
	Macro avg	0.97	**0.98**	**0.98**	**0.98**

Table 9: Performance evaluation of various Deep Learning models trained on the SMS Spam Filtering dataset.

Model	Class	Unbalanced	SMOTE	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1	F1
SVM	Negative	0.00	**0.33**	0.31	0.32	0.32
	Neutral	**0.82**	0.73	0.76	0.77	0.75
	Positive	0.57	0.62	0.61	0.60	**0.63**
	Macro avg	0.46	0.56	0.56	**0.57**	**0.57**
RF	Negative	0.00	0.18	0.17	**0.28**	**0.28**
	Neutral	**0.79**	**0.79**	0.78	0.76	**0.79**
	Positive	0.51	**0.61**	0.57	0.59	0.60
	Macro avg	0.44	0.53	0.51	0.55	**0.56**
LR	Negative	0.08	0.34	**0.36**	0.33	**0.36**
	Neutral	**0.81**	0.76	0.79	0.77	0.78
	Positive	0.60	**0.64**	0.62	0.63	0.60
	Macro avg	0.49	0.58	**0.59**	0.57	0.58

Table 10: Performance evaluation of various Machine Learning models trained on the Sentiment Analysis dataset.

Model	Class	Unbalanced	LMOTE	TLMOTE(M)	TLMOTE(R)
		F1	F1	F1	F1
CNN-LSTM	Negative	0.19	0.26	**0.31**	0.28
	Neutral	**0.78**	0.75	0.77	0.77
	Positive	0.59	**0.60**	0.59	**0.60**
	Macro avg	0.52	0.54	**0.56**	0.55
ACBiLSTM	Negative	0.11	0.20	0.23	**0.25**
	Neutral	0.77	0.77	**0.78**	**0.78**
	Positive	0.55	**0.56**	0.54	**0.56**
	Macro avg	0.48	0.51	0.52	**0.53**
ARC	Negative	0.21	**0.30**	0.23	0.28
	Neutral	0.75	0.71	**0.79**	0.77
	Positive	0.59	0.59	0.59	**0.63**
	Macro avg	0.52	0.53	0.54	**0.56**

Table 11: Performance evaluation of various Deep Learning models trained on the Sentiment Analysis dataset.

Upon dimensionality reduction using Principal Component Analysis, synthetic samples generated using TLMOTE more closely and uniformly resemble the original samples that those generated using LMOTE We observed that synthetic samples from TLMOTE more closely follow the boundary of the original samples in the PCA plots in Fig-

ure 1, as compared to those generated using LMOTE, which are skewed further away from the original data points. Further, the distribution is more uniform in case of TLMOTE samples.

Synthetic datapoints generated using TLMOTE (R) show significantly lower duplicity than those generated using LMOTE for consecutively occurring 5-gram seeds We observed that synthetic samples generated using TLMOTE (R) in Table 6 show substantial variations even for consecutively occurring 5-grams extracted by the topic-based n-gram generator. This reduces the possibility of the classifiers to overfit on the training set, and thus help improve performance on the test set.

4 Conclusion

In this work, we presented Topic-based Language Modelling approach for Text Oversampling (TLMOTE), a variation of LMOTE with three language models with 4, 5 and 6 tokens as input, respectively, used in a randomized order, and 5-gram seeds extracted based on the words relating to the most important topics in the minority class data points. Experimental evaluations show that classifiers trained on datasets oversampled with TLMOTE outperform those trained on datasets oversampled using other approaches like SMOTE and LMOTE in terms of minority class F1 score and macro-averaged F1 score.

References

Almeida, T. A.; Hidalgo, J. M. G.; and Yamakami, A. 2011. Contributions to the study of sms spam filtering: New collection and results. In *Proceedings of the 11th ACM Symposium on Document Engineering*, DocEng '11, 259–262. New York, NY, USA: Association for Computing Machinery.

Anand, A.; Gorde, K.; Antony Moniz, J. R.; Park, N.; Chakraborty, T.; and Chu, B.-T. 2018. Phishing url detection with oversampling based on text generative adversarial networks. In *2018 IEEE International Conference on Big Data (Big Data)*, 1168–1177.

Blei, D. M.; Ng, A. Y.; and Jordan, M. I. 2003. Latent dirichlet allocation. *J. Mach. Learn. Res.* 3(null):993–1022.

Chawla, N.; Bowyer, K.; Hall, L.; and Kegelmeyer, W. 2002. Smote: Synthetic minority over-sampling technique. *J. Artif. Intell. Res. (JAIR)* 16:321–357.

Guo, X.; Zhang, H.; Yang, H.; Xu, L.; and Ye, Z. 2019. A single attention-based combination of cnn and rnn for relation classification. *IEEE Access* 7:12467–12475.

Han, H.; Wang, W.-Y.; and Mao, B.-H. 2005. Borderline-smote: A new over-sampling method in imbalanced data sets learning. In Huang, D.-S.; Zhang, X.-P.; and Huang, G.-B., eds., *Advances in Intelligent Computing*, 878–887. Berlin, Heidelberg: Springer Berlin Heidelberg.

He, H.; Bai, Y.; Garcia, E. A.; and Li, S. 2008. Adasyn: Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322–1328.

Jolliffe, I. 2002. *Principal component analysis*. New York: Springer Verlag.

Kamyab, A.; Liu, G.; and Adjeisah, M. 2021. Attention-based cnn and bi-lstm model based on tf-idf and glove word embedding for sentiment analysis. *Applied Sciences*.

Leekha, M.; Goswami, M.; and Jain, M. 2020. A multi-task approach to open domain suggestion mining using language model for text over-sampling. In *Lecture Notes in Computer Science*. Springer International Publishing. 223–229.

Moreo, A.; Esuli, A.; and Sebastiani, F. 2016. Distributional random oversampling for imbalanced text classification. In *Proceedings of the 39th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '16, 805–808. New York, NY, USA: Association for Computing Machinery.

Negi, S.; Daudert, T.; and Buitelaar, P. 2019. SemEval-2019 task 9: Suggestion mining from online reviews and forums. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, 877–887. Minneapolis, Minnesota, USA: Association for Computational Linguistics.

Pennington, J.; Socher, R.; and Manning, C. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 1532–1543. Doha, Qatar: Association for Computational Linguistics.

Saini, M., and Susan, S. 2019. Data augmentation of minority class with transfer learning for classification of imbalanced breast cancer dataset using inception-v3. In *Pattern Recognition and Image Analysis: 9th Iberian Conference, IbPRIA 2019, Madrid, Spain, July 1–4, 2019, Proceedings, Part I*, 409–420. Berlin, Heidelberg: Springer-Verlag.

Susan, S., and Kumar, A. 2018. Hybrid of intelligent minority oversampling and pso-based intelligent majority undersampling for learning from imbalanced datasets. In Abraham, A.; Cherukuri, A. K.; Melin, P.; and Gandhi, N., eds., *Intelligent Systems Design and Applications - 18th International Conference on Intelligent Systems Design and Applications, ISDA 2018, Vellore, India, December 6-8, 2018, Volume 2*, volume 941 of *Advances in Intelligent Systems and Computing*, 760–769. Springer.

Zhou, C.; Sun, C.; Liu, Z.; and Lau, F. C. M. 2015. A c-lstm neural network for text classification.

v-TCM: Vertical-aware Transformer Click Model for Web Search

Kritarth Bisht
Department of Information Technology
Delhi Technological University
India
kritarthbisht_2k20isy11@dtu.ac.in

Seba Susan
Department of Information Technology
Delhi Technological University
India
seba_406@yahoo.in

ABSTRACT

Understanding [1] and predicting user click behavior on a web search engine results page is critical for online advertising and recommendation engines. The click prediction results can be further used to estimate the relevance of the search engine results. Sequence-to-sequence models are effective for predicting user click behavior since they consider contextual information in the input sequence, and transformers are the recent state of the art in sequence learning. A novel transformer model is proposed in this paper for click prediction and relevance estimation that learns additionally from the vertical information, apart from the query and search engine results that are the inputs for the traditional click models. Vertical information is the representation style of a document in the search engine results page. The proposed vertical-aware transformer click model (v-TCM) takes into account both the position bias and the vertical bias of the search engine results. From our experiments, it is proved that v-TCM outperforms the existing click models for the click prediction and relevance estimation tasks.

CCS CONCEPTS

• **Information systems**→Information retrieval→Evaluation of retrieval results • **Computing methodologies**→Machine learning

KEYWORDS

Click prediction, relevance prediction, transformer, vertical information, web search

1 INTRODUCTION

Click models are the predictive models that are used for predicting user clicks over a browser's results page known as the Search Engine Result's Page (SERP). Earlier, most of the click models were based on Probabilistic Graphical Models (PGM) which is a mathematical way to encode the probability-based relationship between different random variables in the form of a graph. Click models like dynamic bayesian model (DBN) [1], user browsing model (UBM) [2] and dependent click model (DCM) [3] are PGM based. Nowadays, most of the latest click models are based on deep neural networks. Click models like Neural Click Model (NCM) [4] and Click Sequence Model (CSM) [5] are some of the latest models based on the neural network framework that have outperformed the PGM-based models for click prediction [6].

The traditional click models have the query and the search engine results as their inputs. Most of the click models map the relationship between examinations and clicks, evaluating future clicks based on the past observations. There are relatively few works that have explored the incorporation of vertical information of SERP as one of the inputs to the click model. The result page for a search engine contains heterogeneous results federated from verticals of different types such as multimedia, application and text verticals (examples: Image, Video or News), along with the ordinary or organic results, as illustrated in Fig. 1.

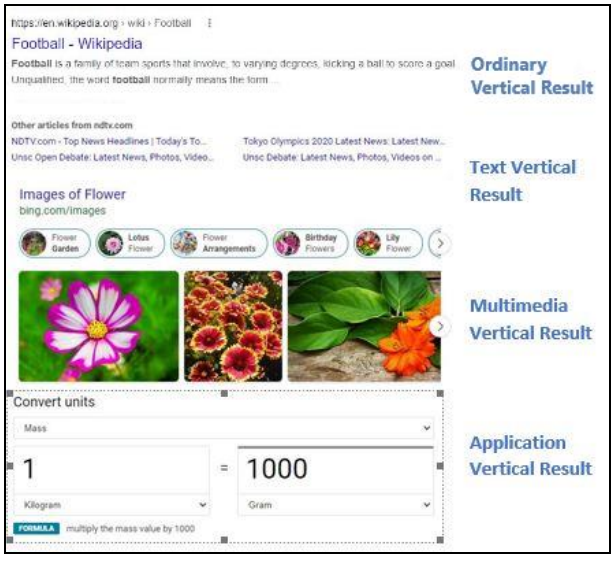


Figure 1: Different types of verticals in SERP

These vertical results seriously influence the user behavior during browsing. Users tend to get attracted by the vertical results and their visual presentation, that may increase the examination probability of nearby results [7]. Examples of some previous click models that have made use of verticals are the vertical-aware click

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s). To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
SAC'22, April 25 –April 29, 2022, Brno, Czech Republic
© 2022 Copyright held by the owner/author(s). 978-1-4503-8713-2/22/04. . . $15.00
DOI: https://doi.org/10.1145/3477314.3507214

model in [8] and the federated click model in [9]. Chen *et al.* [9] were one of the first to model user behavior using vertical results. They found that users are more likely to examine the vertical results and the ordinary web documents nearby. The intuitiveness test by Chuklin *et al.* in [10] established that the document placed near an attractive vertical will have more relevance score. Their work highlights the significance of the placement of verticals for relevance estimation of the search engine results. The vertical-aware model in [8] proposed improvements by incorporating the vertical results, but the relationship between the vertical score and click probability is hard-coded and that may restrain system performance. To shed light on the above-mentioned issue, we have proposed a novel click model that incorporates the vertical/non-Web-page results of the SERP into the input embedding layer of a multi-head attention transformer model [11]. Transformers are the latest state of the art in sequence-to-sequence learning that are rapidly replacing Long Short-Term Memory (LSTM) [12] in the encoder and decoder modules. Some recent works have started adapting transformers for constructing click models. Li *et. al.* (2020) proposed the InterHAt model [13] that uses a transformer with multi-head self-attention for feature learning for the task of click prediction. The distinguishing feature of this model is that it uses attention aggregation to capture high-order feature interactions. Our contribution is new from the perspective that we incorporate the vertical information type of each document in SERP (vertical bias) as an additional input to the encoder of a multi-head self-attention based transformer, apart from the query and the ranked search engine results (position bias) which is the conventional input of click prediction models. The main contributions of our paper are summarized as follows.

- A novel vertical-aware transformer click model called v-TCM is proposed which takes the vertical bias into consideration. The new model is able to incorporate both the position and vertical biases into a calculable framework.
- Transformers with multi-head self-attention constitute the state of the art in sequence learning, and this model is used for learning the position and vertical biases from user logs.
- The model achieves significantly better performance in both click prediction and relevance estimation as compared to existing click models.

The organization of the rest of this paper is as follows. Section 2 describes the proposed model, Section 3 describes the experimental setup and results, and Section 4 draws the final conclusions.

2 VERTICAL-AWARE TRANSFORMER CLICK MODEL

In this section, we describe the architecture of the proposed vertical-aware Transformer Click Model. Click modeling can be interpreted as a sequence modeling task. The query and the search engine results define the input sequence, and the prediction of whether the user will click on a search result or not is recorded for the 10 results displayed on the search page, and this defines the output sequence. Recurrent neural networks, specifically LSTMs, have been used for sequence-to-sequence modeling and prediction [12, 14]. Neural networks are found to capture context well using attention mechanism [15], and that improves the click prediction and relevance estimation results, as proved by researchers in [16]. However, since its inception in 2017 [11], transformers have overtaken LSTMs in most of the sequence modeling tasks. The primary advantage is the parallel computing and multi-head self-attention incorporated in the transformer architecture. In our work we propose to use the transformer for click prediction. We feed the query, position bias (ranked search engine results) and vertical bias as input to the transformer encoder, along with the target clicks. Fig. 2 shows the architecture of the proposed vertical-aware transformer click model that we call as v-TCM.

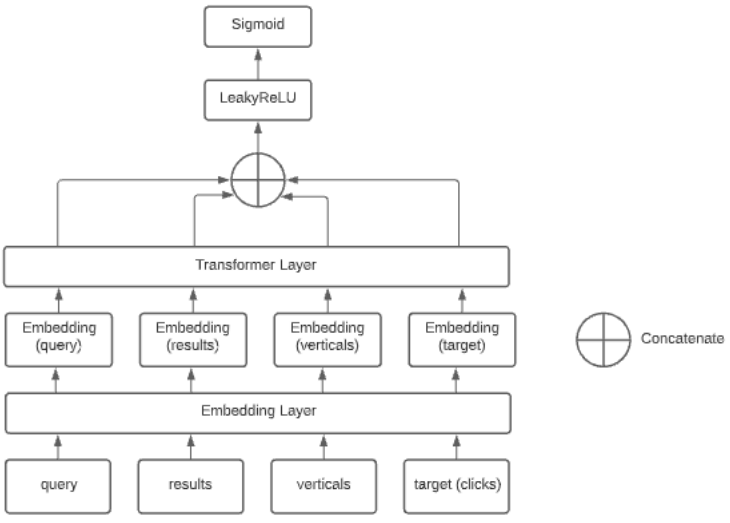


Figure 2: The architecture of the proposed vertical-aware Transformer Click Model (v-TCM)

The transformer PyTorch embedding layer is used to embed all the input features into distributed fixed-size 192-dimensional vectors. The input features are query q, search engine result d_t, and the vertical type v_t of each result, where t=1,2,…,10. We concatenate all the input embeddings and that is given as input to the transformer encoder, along with the embedded target click sequence, as shown in Fig. 2. Our multi-head transformer architecture used four heads whose outcomes are concatenated as shown in Fig. 2. To avoid overfitting, we make use of dropout and Leaky ReLU in both self-attention and feed forward network sub-layers. We then use three fully connected layers to learn the interaction among the dense features. Since the problem is related to whether the user will click on a search result or not, therefore it is a binary classification problem, and thus we use the sigmoid function in the output unit. To train the model, the loss function used is the cross-entropy loss function.

3 EXPERIMENTAL SETUP AND RESULTS

Through our experiments, we aim to answer the following research question: - does the transformer-based click model perform better in terms of click prediction and relevance estimation, and how incorporating the verticals into the click models affects the overall performance. In this section, we first outline the experimental setup. The results obtained are then analyzed with respect to the click prediction accuracy and relevance estimation.

3.1 Dataset

We conducted our experiments on an open log-based session dataset named TianGong-ST [18]. The details are shown in Table 1. The SERP data is refined from the 18-day user log recorded by search engine Sogou.com. The dataset is split at the source into train, test and validation sets. The model is trained on the training set. The hyperparameters were set based on the model's performance on the validation set. Finally for calculating the performance and evaluation metrics, the testing set was used. As noted from Table 1, there are 4491 unique verticals in this dataset that are assigned unique IDs.

Table 1: Dataset statistics

Dataset	TianGong-ST
#sessions	147,155
#unique queries	40,596
#unique doc	263,384
#unique verticals	4491
search engine	Sogou.com
collection	newly crawled in Dec, 2018 (Updated Aug, 2020)

3.2 Performance metrics

We evaluate the performance of v-TCM for click prediction and relevance estimation. For evaluating the click prediction results, we compute the perplexity (PPL) on the test set. The definition of click perplexity (PPL) at the rank r is given by

$$\mathrm{PPL(M)} = 2^{\frac{1}{|S|}\sum_{s\in S}\log_2 P_M(C_r=c_r^{(s)})} \quad (1)$$

where, PPL(M) denotes the total perplexity of the model M. $P_M(C_r=c_r^{(s)})$ denotes the probability of click event $c_r^{(s)}$ having rank r in a query session s. Here, S is the total number of sessions. The final perplexity of the model M is calculated by averaging the perplexities over all positions or ranks. Lower the value of perplexity, the higher will be the quality of the model. For relevance estimation, we use the relevance scores estimated by the click models to re-rank the document list, and calculate the normalized discounted cumulative gain (NDCG) according to the human annotation provided in the TianGong-ST dataset. A higher value of NDCG denotes better relevance estimation by the model.

3.3 Baseline models

We compare the performance of our transformer click prediction model v-TCM with state-of-the-art click models like UBM, DCM, DBN, NCM and CSM. While UBM, DCM, DBN are probabilistic graphical models, NCM and CSM are neural models based on LSTM. We implemented these models in accordance with the guidelines in their original papers [1-5]. We have also included the transformer model (without vertical information) proposed recently for Alibaba recommendation machines [17].

3.4 Hyperparameter settings

For training the proposed models, we use a batch size of 64 and the Adam optimizer. Both the hidden layer size and the embedding dimension is fixed to 64. The initial learning rate and dropout rate is set to 0.001 and 0.5, respectively. For each query, the top 10 documents are considered. Weight decay parameter (λ) is set to 10^{-5} to avoid model overfitting. After every 5 iterations we check the model's validation performance and if it is less than the current validation click perplexity, we update it. The entire coding is done in Pytorch using Google Colab notebook, and all the deep models are trained on the Nvidia K80 GPU.

3.5 Results and discussion

In this sub-section, we analyze the results obtained for the proposed and baseline models, which are summarized in Table 2.

Table 2: Click prediction results

Model	Perplexity	NDCG@10
UBM [2]	1.1793	0.8395
DCM [3]	1.1816	0.8238
DBN [1]	1.1846	0.8346
NCM [4]	1.1758	0.8423
CSM [5]	1.1724	0.8567
Transformer [17]	1.1782	0.8598
v-TCM (Ours)	1.1704	0.8622

The baseline models do not include the vertical information as input features. Table 2 shows the average perplexity score and the NDCG@10 score for v-TCM and the state of the art click models UBM, DCM, DBN, NCM, CSM, and transformer without vertical information. As observed, v-TCM achieves the best scores with the highest NDCG@10 value (0.8622) and the lowest perplexity value (1.1704). Fig. 3 showcases the perplexity scores of the three best performing models in Table 2, namely, NCM, CSM, and v-TCM, over different search result positions ranging from 1-10. As we can infer from Table 2 and Fig. 3, v-TCM performs better as compared to the baseline models due to the lower perplexity values at all search result positions, especially for the top ranked results (ranks 1 to 4).

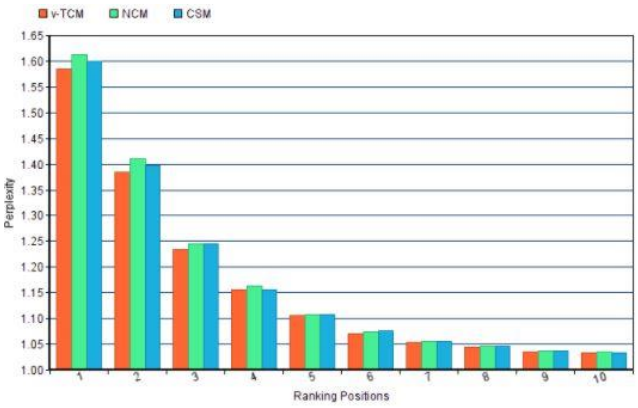


Figure 3: Perplexity scores of NCM, CSM, v-TCM click models for different positions in the SERP

Thus, we can infer that incorporating the verticals into click model improves the performance of the model. The reason is that since the model is taking both position bias and the vertical information into account, this interprets user click behavior better in terms of both perplexity and NDCG@10. Also, the architecture of the multi-head self-attention transformer model improved the effectiveness of our approach.

4 CONCLUSIONS

In this paper, we propose a novel vertical-aware transformer click model (v-TCM) in which we additionally incorporate vertical information as an input to the transformer encoder, apart from the query and search engine results. Experiments are conducted on a public web log session dataset. We find that: 1) transformer model is performing better in terms of perplexity i.e. click prediction, and NDCG@10 i.e. relevance estimation, as compared to the baseline models. 2) incorporating the vertical as an additional input proves fruitful for our cause since it significantly improves the performance, as compared to the baseline models like CSM and NCM that do not take the verticals into account as per their model architecture. More user behavior information must be exploited in the future for improved session and click modeling, like incorporating the vision bias and mouse movement as additional input features in the proposed click model.

REFERENCES

[1] Chapelle, O. and Zhang, Y., 2009, April. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web* (pp. 1-10).

[2] Dupret, G.E. and Piwowarski, B., 2008, July. A user browsing model to predict search engine click data from past observations. In *Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 331-338).

[3] Guo, F., Liu, C. and Wang, Y.M., 2009, February. Efficient multiple-click models in web search. In *Proceedings of the second acm international conference on web search and data mining* (pp. 124-131).

[4] Borisov, A., Markov, I., De Rijke, M. and Serdyukov, P., 2016, April. A neural click model for web search. In *Proceedings of the 25th International Conference on World Wide Web* (pp. 531-541).

[5] Borisov, A., Wardenaar, M., Markov, I. and de Rijke, M., 2018, June. A click sequence model for web search. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval* (pp. 45-54).

[6] Bisht, K. and Susan, S., 2021, May. Weighted Ensemble of Neural and Probabilistic Graphical Models for Click Prediction. In *2021 the 5th International Conference on Information System and Data Mining* (pp. 145-150).

[7] Li, X., Liu, Y., Cai, R. and Ma, S., 2017, February. Investigation of user search behavior while facing heterogeneous search services. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining* (pp. 161-170).

[8] Wang, C., Liu, Y., Zhang, M., Ma, S., Zheng, M., Qian, J. and Zhang, K., 2013, July. Incorporating vertical results into search click models. In *Proceedings of the 36th international ACM SIGIR conference on Research and development in information retrieval* (pp. 503-512).

[9] Chen, D., Chen, W., Wang, H., Chen, Z. and Yang, Q., 2012, February. Beyond ten blue links: enabling user click modeling in federated web search. In *Proceedings of the fifth ACM international conference on Web search and data mining* (pp. 463-472).

[10] Chuklin, A., Zhou, K., Schuth, A., Sietsma, F. and De Rijke, M., 2014, July. Evaluating intuitiveness of vertical-aware click models. In *Proceedings of the 37th international ACM SIGIR conference on Research & development in information retrieval* (pp. 1075-1078).

[11] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł. and Polosukhin, I., 2017. Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).

[12] Bahdanau, D., Cho, K. and Bengio, Y., 2014. Neural machine translation by jointly learning to align and translate. *arXiv preprint arXiv:1409.0473*.

[13] Li, Z., Cheng, W., Chen, Y., Chen, H. and Wang, W., 2020, January. Interpretable click-through rate prediction through hierarchical attention. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 313-321).

[14] Mallick, R., Susan, S., Agrawal, V., Garg, R. and Rawal, P., 2021, March. Context-and sequence-aware convolutional recurrent encoder for neural machine translation. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing* (pp. 853-856).

[15] Goel, R., Vashisht, S., Dhanda, A. and Susan, S., 2021, January. An Empathetic Conversational Agent with Attentional Mechanism. In *2021 International Conference on Computer Communication and Informatics (ICCCI)* (pp. 1-4). IEEE.

[16] Chen, J., Mao, J., Liu, Y., Zhang, M. and Ma, S., 2020, January. A context-aware click model for web search. In *Proceedings of the 13th International Conference on Web Search and Data Mining* (pp. 88-96).

[17] Chen, Q., Zhao, H., Li, W., Huang, P. and Ou, W., 2019, August. Behavior sequence transformer for e-commerce recommendation in alibaba. In *Proceedings of the 1st International Workshop on Deep Learning Practice for High-Dimensional Sparse Data* (pp. 1-4).

[18] Chen, J., Mao, J., Liu, Y., Zhang, M. and Ma, S., 2019, November. TianGong-ST: A new dataset with large-scale refined real-world web search sessions. In *Proceedings of the 28th ACM International Conference on Information and Knowledge Management* (pp. 2485-2488).

International Journal of Research in Engineering, Science and Management
Volume 5, Issue 5, May 2022
https://www.ijresm.com | ISSN (Online): 2581-5792

Weight Optimization and Structural Analysis of an Electric Bus Chassis Frame

Rahul Saini[1], Rishav Jindal[2*], Sandeep Kumar[3], Mohd.Tayyab[4], Ranganath M. Singari[5]
[1,2,3]*B.Tech. Student, Department of Mechanical Engineering, Delhi Technological University, Delhi, India*
[4,5]*Professor, Department of Design, Delhi Technological University, Delhi, India*

***Abstract*: Major cities have seen an increase in motorization in recent decennium, accompanied by deteriorating/poor air quality. Electric buses shows a major obligation in this regard, as these buses have several advantages in respect of lowering noise, pollution, and fuel consumption. Additionally, the costs of keeping an electric bus are around 25% less expensive than those of maintaining a diesel bus, due to the fact that the electric engine does not require the same level of maintenance as a diesel engine. Furthermore, the negative aspect of the electric car is the battery's long-term endurance power. The more energy it takes to run a bus, the heavier it is. It is preferable and more effective to keep the weight of the vehicle elements as low as possible. Our research is focused on the chassis frame, which is an extremely important component of the bus and plays a significant role in supporting the structural safety and strength of the vehicle while also playing an important role in determining the overall weight of the vehicle, which has an impact on the vehicle's performance. This is why we chose the chassis frame as the subject of our investigation. Steel is typically used for the frame since it is sturdy, but it is also heavier and adds weight to the entire bus, diminishing operational effectiveness. The purpose of this post is to illustrate how to reduce the weight of an existing electric bus by replacing its steel components with lightweight aluminium alloys that are both safe and structurally sound while retaining the safety and structural integrity of the frame and chassis. The integrity of the structure is critical in ensuring a safe design. A structural study of the projected chassis frame model will be undertaken using the ANSYS 20.0 simulation programme, using stress and deformation as design constraints, in order to provide a more secure design.**

***Keywords*: Aluminium alloys, ANSYS, Electric bus chassis frame, Structural analysis, Weight reduction.**

1. Introduction

No one can keep up with the fast progress of transportation technology, which is occurring in both developed and developing countries. As a result of the greater movability of people in recent decades, these advancements have become vital. One of the most significant aims of the transportation industry is the creation of transportation networks that allow for the seamless, safe, quick, and cost-effective movement of people from one location to another, among other things. Transportation is such a vital component of human existence that the demand for it is rising, particularly for road transportation, which includes motorcycles, vehicles, buses, and trucks, among other forms of transportation. The need for air transportation is decreasing. Increased demand for autos can let people move more freely, but it can also have severe repercussions for the environment, such as air pollution, if the demand for automobiles increases. The vast majority of internal combustion engines are employed in modern automobiles. It is possible that the exhaust emissions from fossil-fuel driven vehicles will affect the quality of the air in the surrounding area. CO, NOx, CH4, CH4-N, SOx, and other suspended particulate matter (SPM10) are all pollutants that contribute to global warming. CO, NOx, CH4, CH4-N, SOx, and other suspended particulate matter (SPM10). They are all created by motor vehicles and all contribute to global warming in significant amounts, according to the International Energy Agency. When it comes to public transportation, the bus is a cost-effective option that is widely used and has a suitable capacity for passengers. These reasons all lead to a drop in the number of private automobiles on the road in the United States. Diesel engines, which are a type of internal combustion engine, are widely seen in buses and other commercial vehicles. NOx, CO, and other pollutants are emitted from the engine as a consequence of the lack or partial combustion, and these pollutants can pollute the surrounding environment. When diesel engines produce these gases, they also emit black smoke, which is visible when the engines are accelerated or fully loaded. It is also necessary to perform routine maintenance on them in order to keep them running. With the advent of electric buses, these downsides could be mitigated to some extent.

2. Weight Optimization

Obtained by modifying the underlying material or structure, weight optimization can be accomplished with ease. Almost all industries rely on metal for the fabrication of its chassis, which results in a heavy overall product. This forces them to look for new compounds to experiment with. Many researchers have expressed concern about the use of aluminium alloys in light-weight automobile structures [3], [4]. In this study, after looking at aluminium alloys results in ability aluminium alloys especially aluminium 6061-T6 [9], aluminium 7075-T6 [9] with excessive strength-to-weight ratio have been chosen as an alternative to the conventional metal chassis as a consequence of the outcomes in potential of aluminium alloys A description

*Corresponding author: rjindal81462@gmail.com

of those aluminium alloys is presented in the next chapter. The efficiency of machining is evaluated in terms of the willpower required to make the best choice of process adjust factors.

A. Aluminum 6061- T6

It offers yield strength equivalent to mild carbon steel and is one of the most widely used aluminium alloys available. Aluminum 6061 T6 is among the most widely used aluminium alloys available. This alloy, which belongs to the 6xxx group, is primarily made of silicon and magnesium in about the proportions necessary for the formation of magnesium silicide (Mg2Si), which permits it to be heated to exceptionally high temperatures without melting.

The ability of this alloy to generate heat is greatly enhanced by the use of heat therapy. It is well-known for exhibiting acceptable formability, weldability, machinability, and corrosion resistance characteristics. A significant amount of the 6xxx organisation is employed in the construction industry and in other structural applications. As stated in Table 1, the chemical composition of both aluminium 6061-T6 and aluminium 7075-T6 is aluminium 6061-T6 and aluminium 7075-T6 respectively, as indicated in Table 1.

Table 1
Chemical composition

Alloy	Al	Cu	Mn	Mg	Zn	Ti	Cr	Si	Fe
Aluminum 6061-T6	97.8	0.27	0.28	0.86	0.42	0.14	0.069	0.70	0.209
Aluminum 7075-T6	89.7	1.50	0.07	2.40	5.70	0.05	0.190	0.11	0.210

B. Aluminum 7075-T6

Zinc is the most common alloying detail on this kind, and it can be found in concentrations ranging from one to eight percent. The combination of this compound with a reduced percentage of magnesium results in an overall strength improvement of a substantial magnitude. On the whole, the properties of 7075 Al alloy are quite appealing. Low density, ductility, toughness, and fatigue resistance are only a few of the traits that stand out the most. Various additional elements, such as copper and chromium, are only used in trace amounts in the end product on a regular basis and are not used in large quantities.Scandium is also beneficial in small amounts when used sparingly. There are seven different alloys in the 7xxx family, and each is suited for usage in a variety of applications including aircraft structures, cell equipment, and other strongly loaded components. In order to complete this project, a Sunwin SWB6121EV57 electric bus chassis frame constructed of Steel A709M grade 345W has been considered (Structural steel).The rectangular frame was chosen for modelling since it is the most appropriate alternative for big trucks.

The frame's fundamental structural characteristics are as follows:

Table 2
Frame's fundamental structural characteristics

Parameters	Values
Length	11470 mm
Width	2420 mm
Wheel Base	5940 mm
Front Overhang	2550 mm
Rear Overhang	3250 mm

The chassis body is modelled in three dimensions using the Ansys Space Claim software tool. This software programme, which estimates the mass houses of the body, can be used to determine the load on steel and aluminium alloy frames. As it is observed from the outcomes, when steel is substituted with aluminium alloys, the body weight is greatly diminished, as indicated in Table 3.

Table 3

Material	Material density (kg/m^3)	Weight of the frame (kg)	% Reduction in weight
Steel A709M grade 345 W	7850	11,504	–
Aluminum 6061-T6	2700	3957.9	65.61
Aluminum 7075-T6	2800	4103.2	64.33

3. Structural Analysis

With the following step, the goal will be to carry out a structural study of the framework. It is necessary to do structural analysis on the frame since it indicates whether or not it will be able to endure external and internal loads and stresses without cracking or deforming. The finite element method (FEM) is used in order to avoid the large cost and time commitment associated with conducting it on a model, respectively. In order to achieve this, the application Ansys 20.0 is used, which is a general-purpose finite element analysis (FEA) software package.

As shown in Figure 1, the model created using the Ansys Space Claim programme is put into Ansys Mechanical 20.0 for further processing. Following the selection of a material from the materials library, the material is assigned to the model at this point.Since the characteristics of these aluminium alloys were not previously known, they were included in the library as a result of this. This study's frame materials were selected based on their mechanical qualities, which are listed in Table 1 of this document.

Table 4

Properties	Structural steel	Aluminum 6061-T6	Aluminum 7075-T6
Yield strength (MPa)	260	275	505
Ultimate strength (MPa)	450	310	570
Young's modulus (GPa)	200	69.0	71.7
Shear modulus (GPa)	76.9	26.6	26.9
Poisson ratio	0.30	0.33	0.33

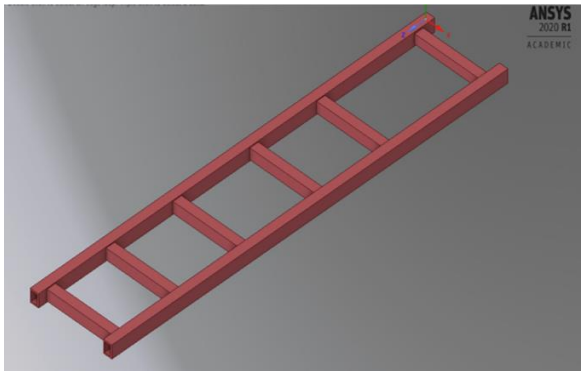


Fig. 1. CAD model of the chassis frame

A. Meshing

Since it dictates how the model should be broken down into finite components, mesh development is a critical stage in the

modelling process. In order to create a continuous mesh, a meshing strategy was devised and implemented. The meshing was accomplished using 2D higher-order tetrahedral components, as seen in the illustration. It is advantageous to utilise this type of element in thin-walled constructions such as the ladder frame since it provides information on local stress. Figure 2 displays a mesh model of the chassis frame with a number of joints, similar to the one in Figure 1.

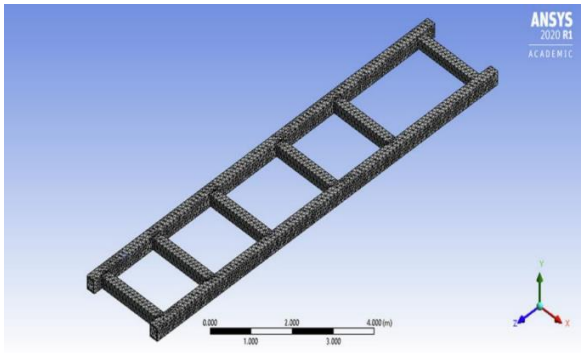


Fig. 2. Meshed model of the chassis frame

B. Loading and boundary conditions

For the purpose of simulating actual vehicle driving situations, it is critical to understand the loads and boundary conditions that will be used in the formulation of the loading parameters and testing parameters. These will be used to replicate driving circumstances including rough roads, aggressive turning, and rapid accelerations, among others. When it comes to model limitations, the structural integrity of vehicle components, as well as the specific loading scenario, are the determining factors. The following is an estimate of the load on the chassis's framework:

Gross vehicle weight (GVW) + payload = 18000 kg.

Capability of the bus = 60 passengers.

Gross weight of the passengers = (60*65) = 3900 kg.

(Assuming avg. weight as 65 kg).

Net weight acting on the frame = 18000 + 3900 = 21900 kg.

Total load acting on the frame = 21900*9.81 = 214839 N.

Those two longitudinal components of the chassis structure are subjected to the greatest amount of stress. Consequently, when a single member is employed, the single member will be subjected to a load equal to one-half (107419.5 N) of the total load operating on the frame as a whole when the frame is not used.

C. FEA results

We picked the design restrictions of stress and deformation as the primary considerations. For the chassis frame produced from three different materials, as indicated in Table 5, the results of the finite element analysis (FEA) for the chassis frame are depicted in a series of figures 3–8.

D. Structural steel frame

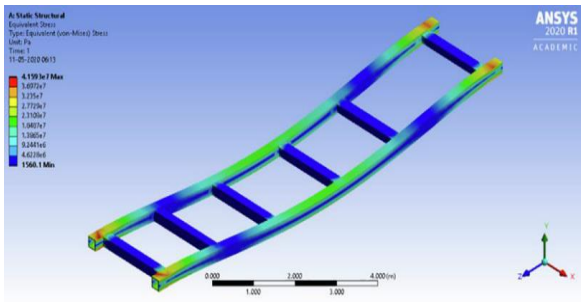


Fig. 3. Equivalent stress of steel frame

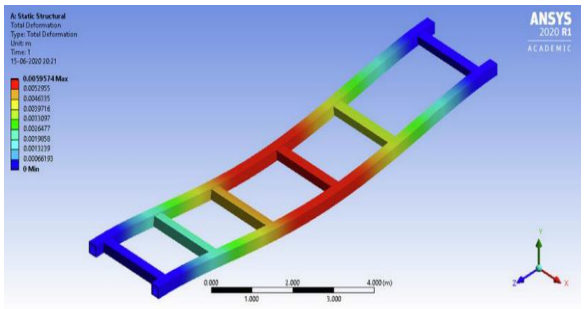


Fig. 4. Total deformation of steel frame

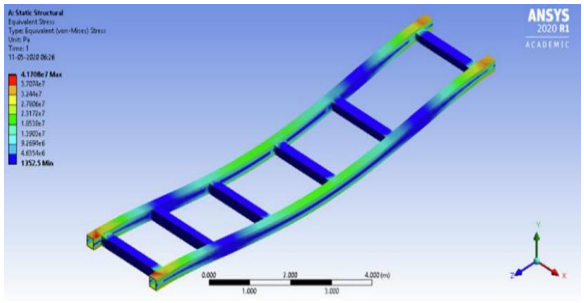


Fig. 5. Equivalent Stress of Alminium 6061-T6 frame

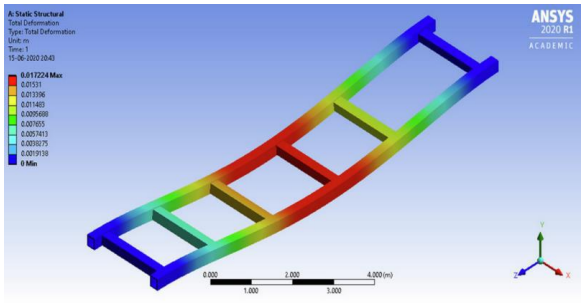


Fig. 6. Total deformation of Aluminium 6061-T6 frame

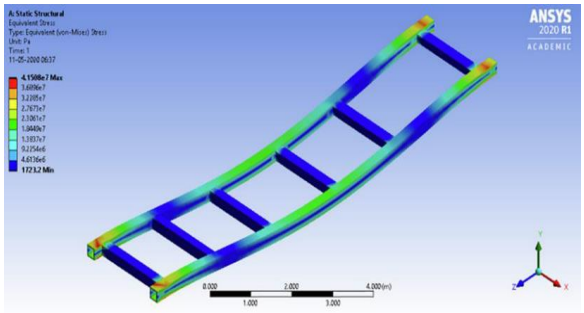


Fig. 7. Equivalent stress of Alminium 7075-T6 frame

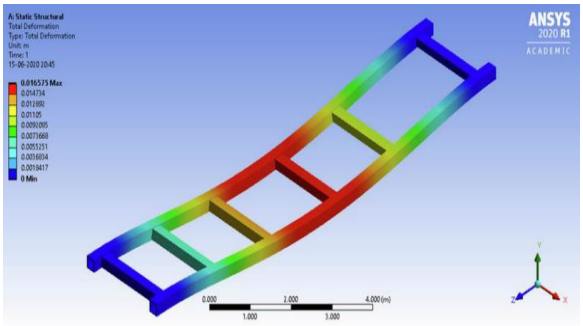


Fig. 8. Total deformation of Alminium 7075-T6

Table 5
Details of the materials

Material	Equivalent Stress (Mpa)	Total Deformation (mm)
Structural Steel	40.56	5.92
Alminium 6061-T6	40.68	17.18
Alminium 7075-T6	40.48	16.52

4. Conclusion

It is possible to lower the weight of the chassis frame by 65.61 % and 64.33 %, correspondingly, by substituting aluminium 6061-T6 and aluminium 7075-T6 for the steel component. The bus's overall performance is significantly improved as a result of the weight reduction achieved. In order to identify safe stress limits for the frame, a finite element analysis was done using finite element software. Upon reviewing the results of the finite element analysis, it was revealed that all three frames were not subjected to stress that exceeded their respective ultimate tensile strengths, proving that the design is safe.

After being subjected to rigorous testing, the aluminium 7075-T6 frame revealed less stress and deformation than its aluminium counterpart, demonstrating that it is a suitable material for the chassis frame, as indicated by the test results.

References

[1] L.E. Teoh, H.L. Khoo, S.Y. Goh, L.M. Chong, Scenario-based electric bus operation: A case study of Putrajaya, Malaysia, Int. J. Transp. Sci. Technol. 7 (1) (2018) 10–25.

[2] Cavazzuti Marco, Andrea Baldini et al. Automotive Chassis Topology Optimization: A Comparison between Spider and Coupé Designs. Proceedings of the World Congress on Engineering 2011, WCE 2011.

[3] M. Tisza, H.S. Lukacs, High strength aluminum alloys in car manufacturing. IOP Conference series: Materials science and Engineering, International Deep Drawing Research Group annual conference, 2018. Volume 418, University of Waterloo, Canada.

[4] Hirsch, Jürgen. "Aluminium Alloys for Automotive Application." Materials Science Forum, vol. 242, Trans Tech Publications, Ltd., Jan. 1997, pp. 33–50.

[5] Lakshmipathy, Jayakumar et al., Reciprocating Wear Behavior of 7075Al/SiC and 6061Al/Al2O3 Composites: A study of Effect of Reinforcement, Stroke and Load, 2014, Tribology in Industry 36.117-126.

[6] Akash Singh Patel, Jaideep Chitransh, Design and Analysis of TATA 2518TC Truck Chassis Frame with Various Cross Sections Using CAE Tools, International Journal of Engineering Sciences and Research Technology.

[7] A. Harikumar, V. Deepanjali, Design and analysis of automobile chassis, Int. J. Eng. Sci. Innov. Technol., vol. 5, no. 1, 2016.

[8] Henry, D. Scott, Fatigue data book – Light structural alloys. ASM International, 2012, ISBN: 978-0-87170-507-5.

[9] R. Joseph Davis, Metals Handbook Desk Edition, ASM International, 1998, ISBN electronic: 978-1-62708-199-3.

[10] S. Rathinave and G. Vigneshkumar, "The Effect of Coolant in Delamination of Basalt Fiber Reinforced Hybrid Composites During Drilling Operation," vol. 3, no. 1, pp. 166-173, 2016.

[11] I. D. Paul, S. M. Sarange, G. P. Bhole and J. R. Chaudhari "Structural Analysis of Truck Chassis Using Finite Element Method" International J. of Multidisc. Research & Advice. In Engg. Vol. 4, No. 1 (January 2012), pp. 85-98, 2012.

[12] Marco Cavazuuti, Luca Splendid, "Structural Optimization of Automotive Chassis: Theory, Setup, Design" Structural and Multidisciplinary Optimization, 2011 pp. 1-3.

[13] K. Chinnaraj, M Sathya Prasad, and C Lakshmana Rao, "Experimental Analysis and Quasi-Static Numerical Idealization of Dynamic Stresses on a Heavy Truck Chassis Frame Assembly" Applied Mechanics and Materials, Vols.13-14, pp. 271-280, 2008.

[14] Untoro Nugroho, Samsudin Anis, RiniKusumawardan Frame analysis of UNNES electric bus chassis construction using finite element method AIP Conference Proceedings 1941, 020017 (2018).

[15] Yang, Zhuo& Deng, Baoqing& Deng, Mengqi & Sun, Gongrui. (2018). A Study on Finite Element Analysis of Electric Bus Frame for Lightweight Design. MATEC Web of Conferences. 175. 03049.

[16] Alvalı, GülüstanTuğçe (2021). Design and Implementation of a Three-Wheel Multi-Purpose Electric Vehicle with Finite Elements Analysis. International Journal of Automotive Science and Technology, vol. 5, 116-125.

[17] Hailemichael Solomon Addisu et al., "Structural Weight and Stiffness Optimization of a Midibus Using the Reinforcement and Response Surface Optimization (RSO) Method in Static Condition," in Modelling and Simulation in Engineering, 2022.

[18] Ismoyo Haryanto et al., "Optimization of Bus Body Frame Structure for Weight Minimizing with Constraint of Natural Frequency using Adaptive Single-Objective Method," International Journal of Sustainable Transportation Technology, vol. 1, no. 1, pp. 9-14, April 2018.

[19] Ninda Kurniadi et al., "Stress Analysis and Size Optimization of Medium Bus Frame using Finite Element Method," in International Journal of Mechanical and Production Engineering Research and Development, vol. 10, no. 4, pp. 53-62, Oct. 2020.

[20] Hari Kishan, A., & Kondalarao, P. (2021). Transient Structural Analysis of Electric Bus Chassis Frame. IOP Conference Series: Materials Science and Engineering, 1185.

[21] Kitchanon Ruangjirakit et al., "Design of Lightweight Electric Bus in Thailand using Composite Materials," 32nd Electric Vehicle Symposium (EVS32) Lyon, France, May 19-22, 2019.

http://dx.doi.org/10.55953/JERA.1104

WIND INTERFERENCE ON A HEXAGONAL SHAPED HIGH-RISE BUILDINGS WITH DIFFERENT OPENINGS

Harsh Goyal[1], Harmanbir Singh[1], Mayank Singh[1] , Ritu Raj[2]

[1] Undergraduate Student, Department of Civil Engineering, Delhi Technological University, Delhi, India

[2] Assistant Professor, Department of Civil Engineering, Delhi Technological University, Delhi, India

[*] Corresponding author rituraj@dtu.ac.in

ABSTRACT:

The wind load and pressure on a tall hexagonal structure altered as the apertures changed in this study. In various conditions, wind angles ranging from 0 to 90 degrees with a 30-degree gap are being used. The distance between the interference building and the model building was gradually raised from 50mm to 200mm at 50mm intervals. At a scale of 1:500, the main building model was studied. The ANSYS CFX package's validity was also tested using a rigid model with dimensions of 100mm x 100mm and a height of 700mm. The k-ε-turbulence model was utilised in the investigation, which was conducted using Computational Fluid Dynamics (CFD) and the ANSYS CFX software package. According to the analysis, changing the percentage of opening area has an influence on building stability because it affects the pressure on the leeward and windward faces. Contours and vortex formation were used to highlight the relationship between different scenarios. As the distance between the interference and model buildings expands, so does the pressure coefficient of the faces. The effects of apertures on the model building were also investigated, with 5-40% openings provided.

Keywords: Hexagonal-Shape Building; ANSYS (CFX); Wind Pressure Coefficient; Contours; Vortex Formation

1. INTRODUCTION:

The introduction of new technology and the rapid rate of scientific research have resulted in an increase in the height, width, and scale of structures, making them even more sensitive to wind damage. As a result, the impact of wind loads on these types of constructions is becoming increasingly important [1,2]. Wind loads on the model structure were initially explored in this work, and then the change in wind loads was investigated when openings were added to our model. Various studies are now being undertaken using computational fluid dynamics (CFD) to examine the effects of wind on tall structures.

http://dx.doi.org/10.55953/JERA.1104

ANSYS (CFX) is a well-known tool for analysing wind load on many types of structures. Buildings with complicated shapes may also be simulated for numerical analysis, and the shape can be partitioned into smaller sections with correct meshing [3,4]. The study [5] looked at the mean pressure coefficient of a tall structure with a 'E' layout both experimentally and computationally. For experimental objectives, wind tunnel testing was employed, while CFD simulation was used for numerical analysis.

The current Indian Standard Code for wind loads over structures contains information on pressure variation and force coefficients. When apertures are provided to irregular structures, however, they fail to manage the complexities of wind loads across them [6]. The primary purpose of this study is to investigate wind loads over isolated structures as well as the effect of interference when a second building with 5-40% openings is introduced during wind testing.

The major goal of this research is to examine wind loads on the buildings with different percentage of openings by comparing the contours and C_P values on different faces. Fig. 1 represents model X i.e., the validation building having a square cross section and dimensions same as that of the hexagonal shaped buildings, the one used for analysis. Height of all the models is 350 metres which is scaled down to 700 millimetres for the ease of simulation. Below Fig.1 shows the buildings of validation model X.

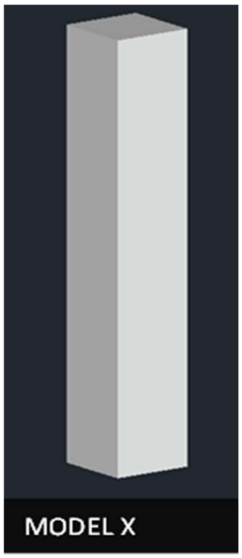


Figure 1: Validation Model

2 ANALYSIS MODEL DEVELOPMENT

2.1 GEOMETRIC DESIGN:

http://dx.doi.org/10.55953/JERA.1104

In the current study, there are three hexagonal shaped buildings named A, B and C with different percentage of openings. These buildings are designed using AutoCAD 3D software and can be seen in Fig.2 Design of the Buildings.

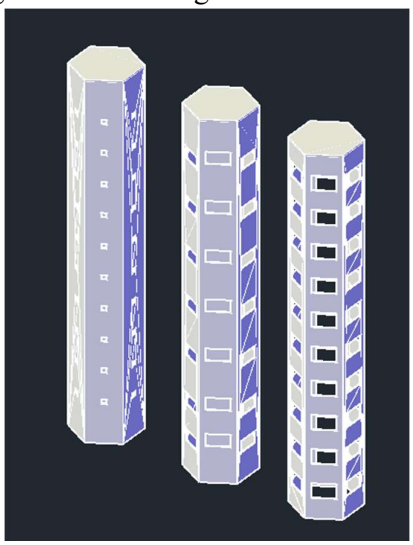


Figure 2: Design of Building A, B and C

The windward face of the building, the one facing the wind is named A while the leeward face, the one on the opposite side is named B. Rest of the faces are named C, D, E and F. The buildings are simulated in a specified domain which is designed as per the guidelines of IS:875 Part -3 [6]. Contours of the Face A and Face B is being compared for different openings along with different wind angles. The wind angles opted for these simulations include 0°, 30°, 60° and 90°.

Description along with the openings percentage of the three buildings are mentioned below –:

1. Model A is the first building with minimum opening percentage of less than five percent of the total area of the building. Total opening area holds up to 12500 mm^2 i.e., 4.97% of the total building area. Figure 3 represents the design of Model A.
2. Model B is the second building with opening percentage between five percent and twenty percent of the total area of the building. Total opening area holds up to 42000 mm^2 i.e., 16.7% of the total building area. Figure 4 represents the design of Model B.
3. Model C is the third building with maximum opening percentage between twenty percent and forty percent of the total area of the building. Total opening area holds up to 72000 mm^2 i.e., 28.66% of the total building area. Figure 5 represents the design of Model C.

http://dx.doi.org/10.55953/JERA.1104

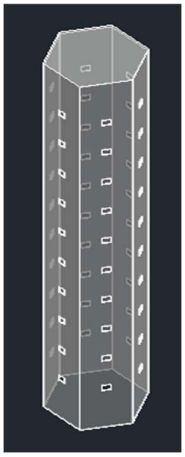


Figure 3
Model A

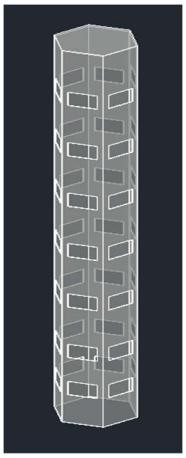


Figure 4
Model B

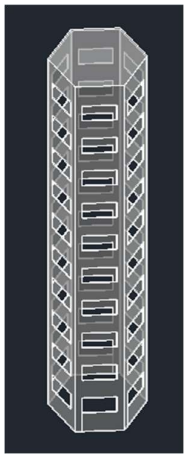


Figure 5
Model C

2.2 ANSYS SIMULATION:

The ANSYS CFX software package is used for simulation, which allows us to apply a variety of boundary conditions. This software's approach consists of five steps: Geometry, Mesh, Setup, Solution, and Result. In the initial stage, we export the analysis building from AutoCAD and specify a domain for the analysis. To avoid flow re-circulation, the domain dimensions are defined in accordance with Indian Standards [6], with the domain boundaries 5h from the model's windward face, lateral faces, and roof and 15h from the model's leeward face (where let h be the height of the model).

Meshing is the following stage, which separates the whole model, including the domain, into tiny tetrahedral nodes of size 0.2m for better results [7]. Inflation is also available to improve the non-slip properties of the building walls. In the current investigation, the model receives 20 inflation layers, as shown in Fig.6 Meshing of Structure.

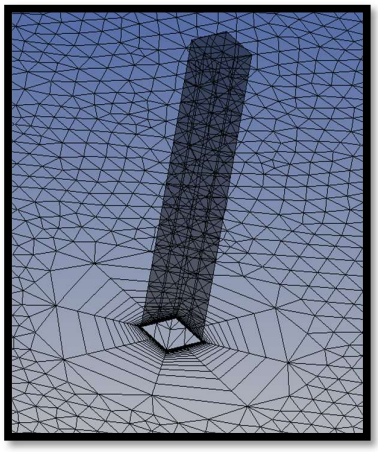


Figure 6: Meshing of Structure

http://dx.doi.org/10.55953/JERA.1104

Next step is setting up the boundary conditions for the whole simulation. The ground and walls of buildings are given a no slip wall condition while the ground and external walls of domain are given a free slip wall condition. For determining the speed of the wind power law is used (Equation 1) . The inputs for the setup are $\alpha = 0.133$, $Z_{ref} = 0.7$ m, and $U_{ref} = 10$ m/s.

$$Power\ law = U_{ref}\left(\frac{Z}{Z_{ref}}\right)^{\alpha} \qquad (1)$$

Then the solution is run which takes approximately two to three hours as hundred iterations are performed on each element. In the final step of result the pressure coefficient is calculated using the Equation 2.

$$C_P = \frac{Pressure}{\left(0.5 * \rho * U_{ref}^2\right)} \qquad (2)$$

Where the value of U_{ref}= 10m/s and the value of density ρ = 1.225 kg/m^3.

The contours for these Cp (Pressure coefficient) values are then plotted and the results are drawn from it. Results along with the conclusions are mentioned in the section 4 and 5 respectively.

3. VALIDATIONS:

Table 1: Comparing Cp values of Validation Model with acceptable Cp values in accordance to IS: 875 Part 3-2015

Pressure Coefficients (C_P)	FACES OF VALIDATION BUILDING			
	A	**B**	**C**	**D**
According to IS:875 (Part III) 2015	0.95	-1.25	-0.7	-0.7
Validation Model	0.78	-0.77	-0.73	-0.71

For validation process an isolated building with square cross section with dimensions same as that of our main observational building is simulated in exactly identical boundary conditions as that of our main simulation. The height of the building is 700mm and the cross-section of the building is 100mm by 100mm throughout the height. Pressure coefficient (C_p) values for all the faces - A, B, C and D (Face A being windward side, Face B being leeward side and Face

http://dx.doi.org/10.55953/JERA.1104

C and D being lateral faces) are compared with the acceptable values as given in IS: 875 (Part III) – 2015 [6].

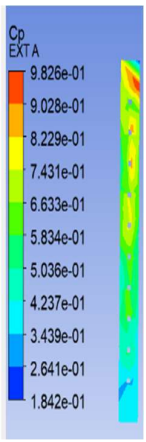
The average error of all four faces was found to be about 12.64 percent which is significantly less and thus, identical boundary conditions was used in the main model.

The conclusions drawn from pressure contours are then discussed below-:

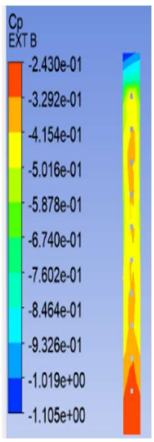
Case 1 Incident wind angle is 0º

Model A

Face A

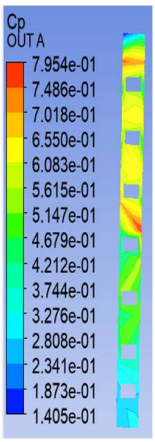


Face B

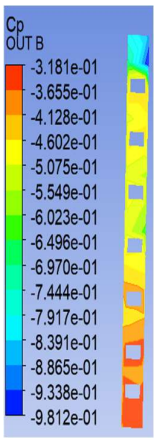


Model B

Face A

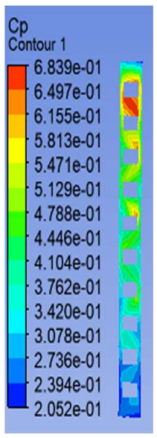


Face B

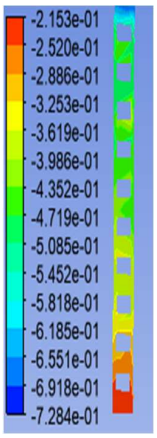


Model C

Face A



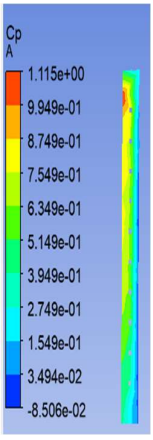
Face B



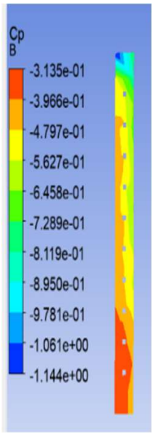
Case 2 Incident wind angle is 30º

Model A

Face A

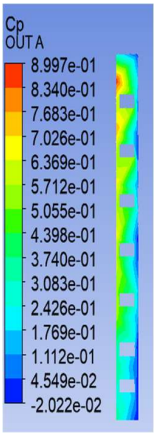


Face B

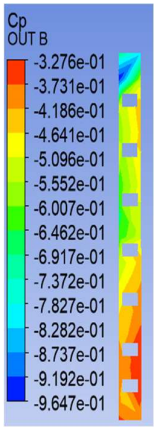


Model B

Face A

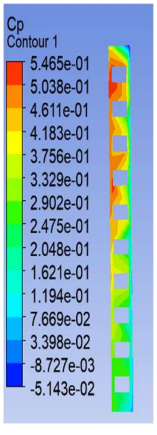


Face B

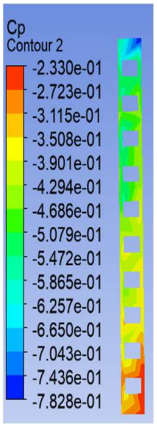


Model C

Face A



Face B



http://dx.doi.org/10.55953/JERA.1104

4.RESULTS AND DISCUSSION:

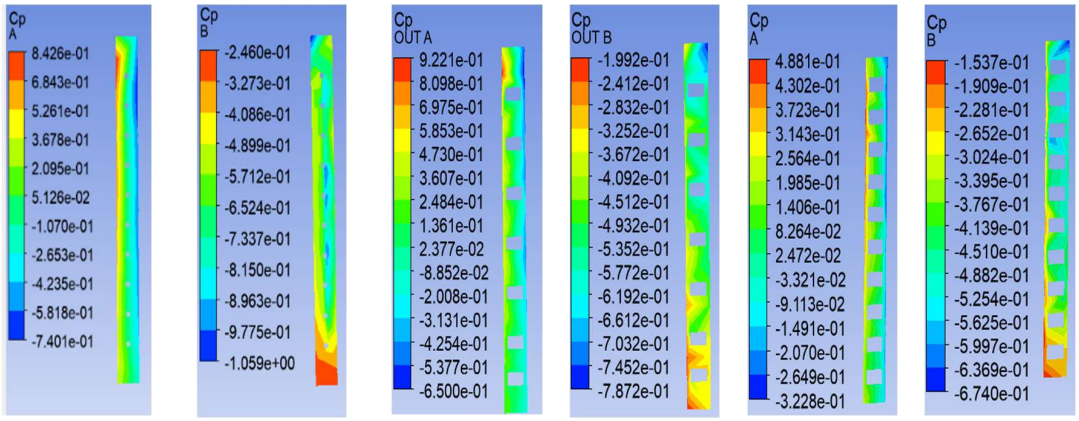
Contours are described as shading on a simulated body in which different colours of shade signify different values. As the shade darkens, so does the associated value. The outlines in this example represent the C_p values. Each contour comes with a caption that shows the related value to the specific colour. The situations are categorised based on the distance between the two structures, which is further subdivided based on wind angles.

The value of C_p increases as the colour darkens. The positive value of the pressure coefficient reflects wind pressure, whereas the negative value represents suction force. A legend is always provided on the side to give us with the values that correspond to different colours so that we can determine which colour corresponds to which value [8,9].

Case 3 Incident wind angle is 60°

Model A **Model B** **Model C**

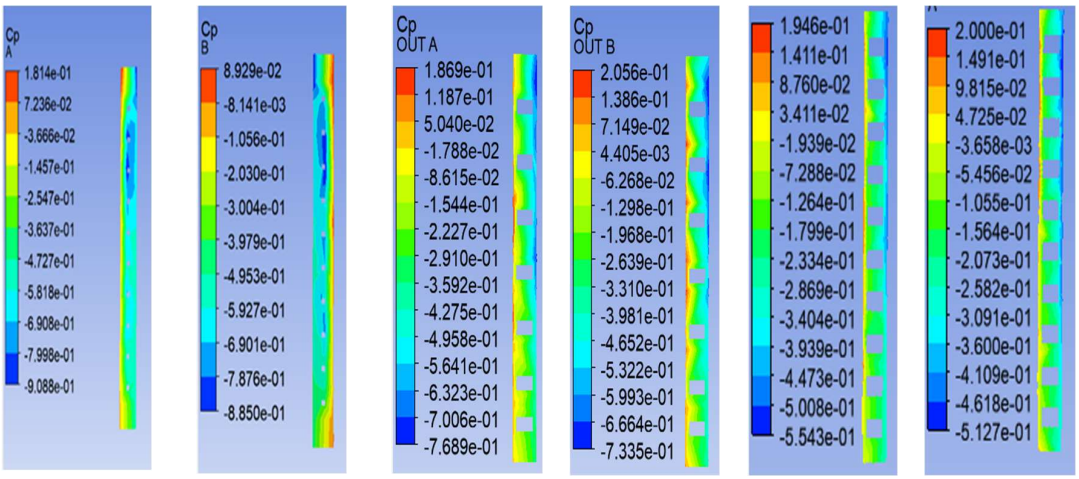
Face A Face B Face A Face B Face A Face B



Case 4 Incident wind angle is 90°

Model A **Model B** **Model C**

Face A Face B Face A Face B Face A Face B



The contours given lead to the following conclusions:

http://dx.doi.org/10.55953/JERA.1104

- It can be seen that range of C_P for face A is decreasing as the opening percentage increases but as the angle of incidence increases the same trend is not observed. For 0 and 30 degrees the same trend is observed while as the angle changes to 60 degrees the C_P range of the model B becomes greater than the other two cases. While for the case of 90 degrees the range of model C in greatest among all.
- 1.115 is the maximum C_p value observed for face A which was found for model A when the wind angle incidence is 30 degrees while the most negative value observed for the same is -0.908 for model A when the incidence angle is 90 degrees.
- For face B it is observed that as the opening percentage of windows increases, the range of the C_p values become less negative as it is observed for the wind incidence angle 0 degrees, the value of C_p for model A, B and C is -1.1, -0.98 and -0.72 respectively. It is observed that in the fourth case where the incidence angle is 90 degrees the range of C_p values become positive in all the three models.
- The reason for the changes in C_p value is because of the formation of vortex i.e., the swirling of wind particles about a particular axis because of which the C_p value decreases for the face A [10,11].

5. CONCLUSIONS:

Wind being a fluid has very complex nature and thus gives rise to many unpredictable flow conditions which needed to be studied in order to build more high-rise structures. The followings conclusions are drawn from this study.

- This study account for the effect of different percentages of openings on high rise hexagonal shaped buildings.
- The validation section discusses about the C_P values obtained for the model X and they were in agreement with those given by IS: 875 (Part III) – 2015. Therefore, Model A is also validated by consequence.
- The study concludes that changing in the percentage of opening area has an effect of building stability as it changes the pressure on leeward and wind ward faces.
- Vortex formation also takes place as there is change in the opening percentage for the high rise building which is a very important and crucial element for the stability of high rise buildings.

REFERENCES:

1. Gu M, Quan Y. Across-wind loads and effects of super-tall buildings and structures. Science China Technological Sciences. 2011 Oct;54(10):2531-41.
2. Raj R, Ahuja AK. Wind loads on cross shape tall buildings. J. Acad. Ind. Res. 2013 Jul;2(2):111-3.

http://dx.doi.org/10.55953/JERA.1104

3. Kumar A, Raj R. Study of Pressure Distribution on an Irregular Octagonal Plan Oval-Shape Building Using CFD. Civil Engineering Journal. 2021 Oct 1;7(10):1787-805.
4. Nagar1a SK, Raj R, Dev2b N. Experimental study of wind-induced pressures on tall buildings of different shapes. Wind and Structures. 2020;31(5):441-53.
5. Bhattacharyya B, Dalui SK. Investigation of mean wind pressures on 'E'plan shaped tall building. Wind and structures. 2018 Feb 1;26(2):99-114.
6. "IS (Indian Standard). 2015. Indian standard code of practice for design wind load on building and structures. IS 875 (Part 3). New Delhi, India: IS."
7. Raj R, Rukhaiyar A, Jayant B, Dahiya K. CFD Analysis for wind flow characteristics of varying cross section tall building using ANSYS.
8. Goyal PK, Kumari S, Singh S, Saroj RK, Meena RK, Raj R. Numerical Study of Wind Loads on Y Plan-Shaped Tall Building Using CFD. Civil Engineering Journal. 2022 Feb 1;8(2):263-77.
9. Meena RK, Raj R, Anbukumar S. Numerical Investigation of Wind Load on Side Ratio of High-Rise Buildings. InAdvances in Construction Materials and Sustainable Environment 2022 (pp. 937-951). Springer, Singapore.
10. Pal S, Raj R, Anbukumar S. Bilateral interference of wind loads induced on duplicate building models of various shapes. Latin American Journal of Solids and Structures. 2021 Jul 26;18.
11. Gaur N, Raj R, Goyal PK. Interference effect on corner-configured structures with variable geometry and blockage configurations under wind loads using CFD. Asian Journal of Civil Engineering. 2021 Dec;22(8):1607-23.

Received: 05 April 2022

Revised: 20 April 2022

Accepted: 25 April 2022

Journal of Engineering Research and Application