Volume 49 • No 3 • July-September 2026
Periodicity: Quarterly • Subject: Multidiciplinary

ISSN 0971-3034

THE INDIAN JOURNAL OF TECHNICAL EDUCATION

Published by
INDIAN SOCIETY FOR TECHNICAL EDUCATION
Near Katwaria Sarai, Shaheed Jeet Singh Marg,
New Delhi - 110 016

INDIAN JOURNAL OF TECHNICAL EDUCATION

Volume 49 • No 3 • July-September 2026

Editorial Advisory Committee

Prof. Pratapsinh K. Desai - Chairman
President, ISTE

Prof. N. R. Shetty
Former President, ISTE, New Delhi

Prof. (Dr.) Buta Singh Sidhu
Former Vice Chancellor, Maharaja Ranjit Singh Punjab Technical University, Bathinda

Prof. G. Ranga Janardhana
Former Vice Chancellor
JNTU Anantapur, Ananthapuramu

Prof. D. N. Reddy
Former Chairman
Recruitment & Assessment Centre
DRDO, Ministry of Defence, Govt. of India
New Delhi

Prof G. D. Yadav
Vice Chancellor
Institute of Chemical Technology, Mumbai

Dr. Akshai Aggarwal
Former Vice Chancellor
Gujarat Technological University, Gandhinagar

Prof. M. S. Palanichamy
Former Vice Chancellor
Tamil Nadu Open University, Chennai

Prof Amiya Kumar Rath
Vice Chancellor, BPUT
Rourkela

Prof Raghu B Korrapati
Fulbright Scholar & Senior Professor
Walden University, USA & Former Commissioner for Higher Education, USA

Editorial Board

Dr. Vivek B. Kamat
Director of Technical Education
Government of Goa, Goa

Dr. Ishrat Meera Mirzana
Professor, MED, & Director, RDC
Muffakham Jah College of Engineering and Technology
Hyderabad, Telangana

Prof. (Dr.) CH V K N S N Moorthy
Director R&D
Vasavi College of Engineering
Hyderabad, Telangana

Prof. C. C. Handa
Professor & Head, Dept. of Mech.Engg.
KDK College of Engineering, Nagpur

Prof. (Dr.) Bijaya Panigrahi
Dept. Electrical Engineering
Indian Institute of Technology, Delhi
New Delhi

Prof. Y. Vrushabhendrappa
Director
Bapuji Institute of Engg. & Technology, Davangere

Dr. Anant I Dhatrak
Associate Professor, Civil Engineering Department, Government College of Engineering, Amravati, Maharashtra

Dr Sunil J Raykar
HoD Mechanical Engineering
DY Patil College of Engg. and Tech.

Dr. Jyoti Sekhar Banerjee
Associate Editor

Dr. Rajeshree D. Raut
Associate Editor

Dr. Y. R. M. Rao
Editor-in-Chief

Copyright (c) Indian Society for Technical Education, The Journal articles or any part of it may not be reproduced in any form without the written permission of the Publisher.

INDIAN JOURNAL OF TECHNICAL EDUCATION

Published by
INDIAN SOCIETY FOR TECHNICAL EDUCATION
Near Katwaria Sarai, Shaheed Jeet Singh Marg
New Delhi - 110 016

INDIAN JOURNAL OF TECHNICAL EDUCATION

Volume 49 • No 3 • July-September 2026

Periodicity : Quarterly

Subscription Rates with Effect from 01 Jan 2024

FOR ONE/TWO YEARS or 4/8 ISSUES

		One year	Two Years
ISTE Life Members	:	Rs. 2000	Rs. 3500
Institutional members of ISTE	:	Rs. 3000	Rs. 5500
Non-member educational & Research institutions and other individuals	:	Rs. 5000	Rs. 9000
Industry/Government Department/ Other organisations.	:	Rs. 6000	Rs. 11000

Note :

1. Above mentioned rates are **exclusive of GST** as applicable.
2. Above mentioned subscription rates are for 04 issues per year only. If any changes in periodicity, rates will differ and communicated to the subscriber.
3. Send the subscription amount by NEFT to the following bank details
   Indian Society for Technical Education, New Delhi
   ISTE Training & Development A/C
   Bank: HDFC Bank
   A/C No. : 50100771116575
   IFSC: HDFC0002055
   Branch: Katwaria Sarai, New Delhi

For details Contact:

The Executive Secretary
Indian Society for Technical Education
Near Katwaria Sarai, Shaheed Jeet Singh Marg
New Delhi -110 016
E-mail : editor@isteonline.org
Website : www.isteonline.in

Note : The Editorial Board does not assume any responsibility for the individual opinions expressed by the authors, and are in no way those of the editor, the institution to which the author belongs or the publisher.

Editorial

Beyond the Ph.D. Thesis: The conventional Ph.D. system evaluates scholars through a comprehensive thesis demonstrating original research, critical analysis, and contribution to knowledge, followed by external evaluation and an oral defence. This approach remains fundamental to academic research. However, in engineering and technology, significant innovations may also emerge through industrial solutions, technological developments, patents, and practical applications. Recognizing such achievements could strengthen the connection between doctoral education and real-world engineering challenges.

China has introduced an alternative route for professional engineering doctoral candidates, permitting them to seek a degree through approved practice-oriented achievements instead of the conventional dissertation. This initiative aims to strengthen university-industry collaboration and promote technological innovation. Eligible achievements may include innovative products, engineering designs, process improvements, and solutions to complex industrial problems. The approach recognizes that substantial engineering contributions can emerge from laboratories, industries, and major infrastructure projects.

Importantly, this model does not eliminate research or documentation. Candidates must demonstrate originality, individual contribution, practical significance, and innovation. Their applications require tangible deliverables, a comprehensive written report, expert assessment, and a formal defence. This provides a valuable lesson for India: doctoral assessment can recognize diverse forms of achievement while maintaining rigorous academic standards. However, China's framework specifically concerns professional engineering doctorates and should not be interpreted as abolishing conventional Ph.D. requirements.

India could explore a regulated practice-based pathway alongside the conventional Ph.D., subject to appropriate regulatory approval. Best practices should include transparent eligibility criteria, independent academic and industry evaluation, verification of practical outcomes, plagiarism screening, and clear intellectual-property safeguards. Publication of original research papers in highly reputed, peer-reviewed journals should also be compulsory, with genuine scholarly contribution prioritized over publication numbers. Such publications invite international scrutiny, constructive criticism, and opportunities to improve research quality. A carefully regulated framework could promote innovation, industry collaboration, and technology transfer without compromising doctoral credibility.

New Delhi **Editor-in-Chief**

31st October 2026

Indian Society for Technical Education

Official E- Mail ID's

SN	E-Mail ID	Purpose
1	**info@isteonline.org**	All General enquires.
2	**istedhq@isteonlione.org**	All official communication.
3	**exesecretary@isteonline.org**	Matter related to Executive Secretary; Policy related & etc.
4	**accounts@isteonline.org**	All matter related to finance and account section.
5	**membership@isteonline.org**	All matters related to, 1. Institute Membership, 2. Life Membership, 3. Student Membership, 4. Faculty Chapter, 5. Student Chapter, 6. Section Share, 7. Student Share & etc.
6	**guesthouse@isteonline.org**	For Guest House Booking.
7	**executive_tech@isteonline.org**	All enquiries related to ISTE official Website, E-Mail ID & etc.
8	**technical@isteonline.org**	For Faculty Development Program (FDP), STTP's, Enquiry related to Conventions, Conferences & etc.
9	**editor@isteonline.org**	Enquiry related to Indian Journal of Technical Education (IJTE) & Submission of Manuscript & etc.
10	**newsletter@isteonline.org**	All matters related to news items in newsletter.

Shaheed Jeet Singh Marg, Near Katwaria Sarai, Opp. Sanskrit Vidyapeeth, New Delhi-110016
istedhq@isteonline.org www.isteonline.in

Contents

1.	**Urban Residential Property Values: Identifying the Key Influencing Parameters** Sneh Fandian, Manoj Panwar	**1**
2.	**Optimisation of Biogas Production from Co-Digestion of Organic Waste Streams of Kolhapur Municipal Corporation Using Response Surface Methodology** Prakash Lagamavar, Sudhir B. Desai, Mrunalini P. Jagtap	**11**
3.	**Design Chart Development for Sustainable Sheet Pile Retaining Structures** Ramya L, Padmavathi M	**20**
4.	**Feasibility of Collaborative Robot-Assisted Wall Laying Using Interlocking Compressed Earth Blocks** Vishal Kumar Chaudhary, Hemant Kumar Prajapati	**30**
5.	**Similitude Framework for Scale Modelling of Dry-Stacked Interlocking Compressed Earth Block Walls** Vishal Kumar Chaudhary, Sneha Kumari, Sumit Suman	**39**
6.	**Comparative Study of Autoclaved Aerated Concrete (AAC) Blocks and Conventional Concrete Blocks for Sustainable Construction** Abhishek Kumar, Ravi Ranjan Kumar, Ishant Kumar, Prabhakar Kumar	**49**
7.	**Identifying the Best Interpolation Technique for $PM_{2.5}$ Concentrations in Delhi** Sumit Suman	**60**
8.	**A Preference-Aligned Vision Transformer Framework for Camera-Based Safety Monitoring in Building Construction** Vishal Kumar Chaudhary, Pankaj Kumar	**68**
9.	**Cross-City Transfer Learning for Daily Rainfall Forecasting at Pune, India** Sumit Suman, Vishal Kumar Chaudhary, Tanmay Kumar	**78**
10.	**The Role of Augmented and Virtual Reality in Electrical Engineering: From Education to Industry 5.0.** Malliga P	**84**
11.	**X-DRIFT: A Dual-Stage Mutual Information Pruning Framework for Explanation Stability in Edge-IoT Intrusion Detection Systems** Minakshi Panwar	**94**
12.	**Toward Adversarial Robust and Edge-Deployable Network Intrusion Detection: An Attention-Augmented Architecture with Data Validation** Narasimha Swamy Kandelli, K. Gangadhara Rao, Simhadri Mallikarjuna Rao	**105**

13. **Secure Data Protection and Retrieval for an Individual Including Comprehensive Data through Aadhar** 111
G. Wiselin Jiji, Jasper Gifty Vijayan, J. Jinolin

14. **ResNet50 Based Sign-to-Text-to-Speech Translation** 118
Sheetal A. Takale, Vaishnavi D. Jankar

15. **Dynamic Wireless Electric Vehicle Charging with Seamless LCC-S Pad Handover Control** 129
Suganthi. J.

16. **Comparative Analysis and Genetic Algorithm Optimization of Multi-Transistor FinFET SRAM Topologies for SNM, Leakage, and Delay Improvement at 18nm** 138
Govinda M R, Mallikarjun Sarsamba, Parameshwara M. C.

17. **Bending analysis of Compact geometry Monopole Antenna for Wearable Applications** 152
Ambresh P. A.

18. **Massification of Technical Education in India: Implications for Quality and Academia–Industry Alignment** 158
Keshav Kant

19. **Mechanical Durability of Contemporary Dental Materials: An Engineering Review and Research Roadmap** 168
Amol B. Patil, Amit S. Shelake, Jitendra B. Patil

20. **Manufacturing-Structure-Performance Engineering of Advanced Dental Materials: Additive Manufacturing, Surface Engineering and Computational Design** 190
Jitendra B. Patil, Amol B. Patil, Yogesh S. Andhale

21. **Human Resource Management Practices and Their Influence on Service Excellence in Luxury Hotels: Evidence from Chennai City** 213
Balraj V N, G. Anitha

22. **An Empirical Investigation of Human Resource Management Practices and Service Excellence in Luxury Hotels** 222
Balraj V N, G. Anitha

23. **Exploring the Determinants of Tourism and Hospitality Career Choice among Undergraduate Students: A Study of South Indian Universities** 230
Mathew Stephenson J, Ahamed T Z

24. Career Attitudes, Employability Perceptions, and Industry Expectations among Tourism and Hospitality Undergraduates in South India **238**
Mathew Stephenson J, Ahamed T Z

25. Assessing Service Quality and Customer Experience in Premium and Luxury Hotels: Evidence from Chennai City **248**
Michael Santhosh A, R. Kannan

26. A Comparative Study on the Quality of Work Life of Male and Female Nurses in Kerala Hospitals **256**
Prasad P, Shyni. P, G. Manimekala

27. Traditional Cuisine as a Tourism Resource: An Assessment of Southern Districts of Tamilnadu **263**
S. Sugasri, Ahamed T. Z.

28. Human Resource Management Practices and Employee Performance in Luxury Hotels: Implications for Tourism Development in Madurai **270**
Hamasadhwani S L, Ahamed T Z

DOI - 10.68648/ijte-q-2026-49-03-001

Urban Residential Property Values: Identifying the Key Influencing Parameters

Sneh Fandian
Assistant Professor
DCRUST, Murthal,
Sonipat, Haryana
✉ snehfandian.arch@dcrustm.org

Manoj Panwar
Assistant Professor
DCRUST, Murthal,
Sonipat, Haryana
✉ manojpanwar.arch@dcrustm.org

ABSTRACT

Property value is a dynamic concept influenced by various parameters. The limited number of controlling parameters for property value is discussed in the existing literature, highlighting the need for a comprehensive list of factors and their relative importance. This paper focuses on identifying factors that affect property values in urban areas and provides their rankings and weights. A systematic review of literature was conducted following the PRISMA model by searching for terms like "factors impacting property values," "valuation factors of real estate," and "parameters of property values" across Google Scholar, ResearchGate, and Scopus. A total of 104 factors were identified through an exhaustive survey. Using a two-cycle Delphi method with a 5-point Likert scale, the list was refined to 67 and then further reduced to 37 factors under 10 major categories, which are further categorized into three levels: Macro, Meso, and Micro. These factors and sub-factors were ranked and weighted using the relative ranking method. Experts from academia and industry participated in the Delphi process. Location, connectivity, and accessibility were found to have the most tremendous impact on property values. This study has identified and categorized factors, regardless of geographic limitations, using PRISMA techniques and a relative ranking approach.

KEYWORDS *: Property values; Real Estate; Parameters; Accessibility.*

INTRODUCTION

Property value is a crucial element in both domestic and national economies. A property is a multi-dimensional entity [1]. These factors and their impact vary depending on the place and city (Wyatt, 1997) [2]. Therefore, the identification of critical factors of property values is essential for all stakeholders [3]. Property values increase with the availability and quality of physical and social infrastructure. [4]. A buyer might be interested in paying a substantial amount for real estate, based on the availability of basic amenities, including access to utilities such as water supply and electricity [5]. It has been shown that cities with adequate transportation and economic growth have consistently seen increases in land and property values [6]. Infrastructure availability is highly important for property values[7]. To estimate a property, the well-documented influence of neighborhood services, property features, and amenities on property values was demonstrated. Studies conducted in various geographical locations emphasize the significance of location and the stage of urban development [8].

It has been found that previous individual studies have emphasized limited parameters, and expert opinions have also been utilized in prior research to evaluate the factors that affect property values and location-specific factors. Furthermore, relative ranking and weights were not used in property management studies to choose variables.

So, the primary focus of the present investigation is to identify the main attributes that affect property values in a region by determining the relative weights and rankings. The necessary data for this research were collected through a systematic literature review, which was further refined by a two-cycle expert survey using a 5-point Likert scale along with a relative ranking

method. This research focuses solely on the factors that affect residential homes. The results will help those working in urban planning, real estate development, and property management understand the most significant variables that affect their decisions. The methodology can also be implemented in similar contexts.

METHODOLOGY

The adopted research approach, as shown in Table 1, consisted of a systematic literature review using the well-established PRISMA technique, as well as a two-cycle Delphi survey. A systematic review was done through a comprehensive search of publications related to the value of property factors.

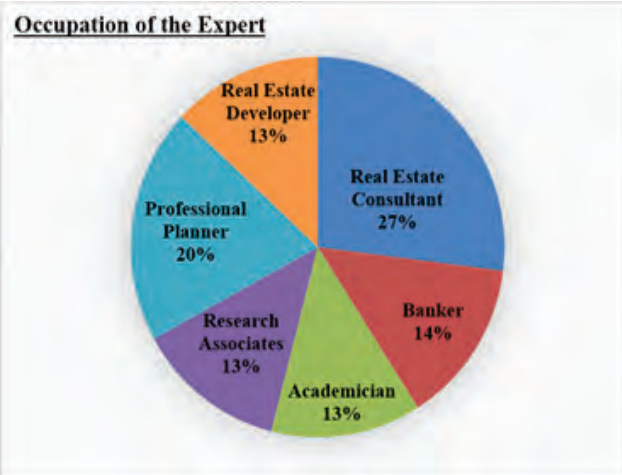
Table 1. Methodology adopted for the research

Step 1: Systematic Literature Review Via PRISMA Technique Identification of 104 factors that affect property values with keywords: Property values, Real Estate Pricing, Land Pricing, and Infrastructure. Reduction of 104 factors into 67 factors.
Step 2: First Cycle Delphi Survey based on 5 points Likert Scale 67 factors were reduced to 37 factors and further classified into 10 categories
Step 3: 2nd Cycle Delphi Survey based on a 5-point Likert scale The 10 main factors are further grouped into 3 main categories – meso, micro and macro. Relative Ranking and weight of 37 factors within the categories within the sub-factors.

Various papers and articles were searched using keywords such as property values, real estate pricing, land pricing, and infrastructure with Boolean operators. With a time frame filter of 1990-2024, almost 200 papers were downloaded from Google Scholar, ResearchGate, and Scopus. Of these, only 65 were relevant to the study.

After a thorough review, 104 factors were identified through the literature review, with redundancy, which was considered improper as discussed, and subsequently filtered down to 67 factors. In the first cycle of the Delphi survey, experts were asked to rank the 67 factors. Following the first cycle, 19 components were eliminated because they were either combined with other elements or designated as important factors. In the second cycle of the Delphi survey, experts were asked to re-rank the factors selected in the first cycle of the survey. This involved identifying the 37 key factors that affect property values. A questionnaire-based survey employing a 5-point Likert scale was designed to gather two-cycle expert opinions on the factors that were shortlisted and categorized. One was the least significant, and five was the most significant. Those 37 factors were grouped under ten major headings of attributes. Then, the 10 main variables were grouped into three categories: micro, meso, and macro, based on the scale of the influence of the attributes. These 37 factors were further examined in detail to understand their impact on property values and the direction of causality.

The most important decision factor in a Delphi survey is the selection of experts. [9][10]. For the research, the experts include academics, real estate professionals, urban and transportation planners, Real Estate developers, and other professionals with at least 10 years of experience, including real estate bankers and real estate investment professionals.



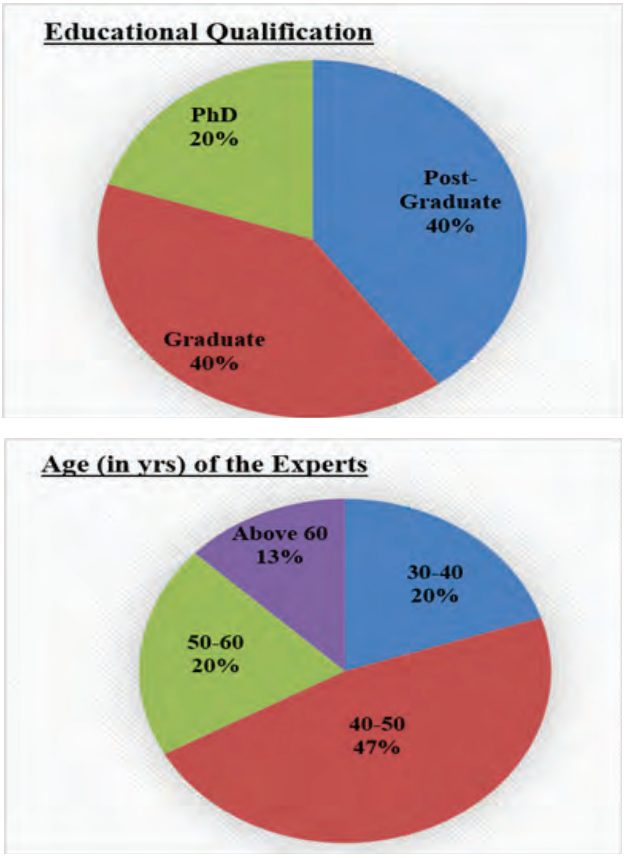


Fig 1: Professional details of the Experts

In a two-cycle Delphi survey, a questionnaire-based survey was conducted by 15 experts from the fields of property professionals and urban planners using Google's survey form. The professional details of the experts are presented in Fig 1. Out of 15 experts, 4 were Real Estate Consultants, 2 were investment bankers, 2 were academicians, research associates, and real estate developers, and 3 were professional urban/town planners.

LITERATURE REVIEW

The value of real estate is a crucial aspect of the market that depends on several factors. Knowing these things and how they relate to each other is a significant part of determining a property's value [3]. Location, physical attributes, legal and environmental factors, accessibility, and proximity to infrastructure are key elements that affect property values [4]. The investigation examined the impact of MRTS and BRTS on property prices in the housing sector. It concluded that property values were greater in controlled areas than in the catchment regions, taking into account neighbourhood features. The study examined the effect of MRTS on housing pricing in the pre- and post-construction stages of the metro corridor and divided the factors into Spatial and Temporal Categories. Spatial attributes took into account the characteristics of spatial dimensions. Temporal attributes are considered the dimension of time [11].

The study identified 80 parameters from various literature studies and shortlisted the factors into 25 after discussion with experts in real estate. The study examined the influence of structural and environmental factors on real estate values in Artivin City. It concluded that the structural value was five times greater than the environmental value. Other research also identified accessibility to social and physical amenities as an important parameter that affects property values in any location. The study by Anantsuksomsri & Tontisirin determined how mass transit directly influences the residential property values in the Bangkok Metropolitan Region, Thailand. Current Research concluded that proximity to MRTS positively influences residential property values. A study by M. Zhang & Wang 2013 also concluded that the property values are higher in the case of MRTS, and values go down as we move away from the transit station. It emphasized the notable influence of the Mass transit system on property and compared transit technologies, such as BRTS, LRTS, and MRTS, in Beijing. The research concluded that urban growth and land values are directly proportional, as the land value increases with an increase in urban sprawl, and emphasized the proper future planning of cities and urban sprawl [13]. The study emphasized the other factors, such as location, economy, land use type, urban sprawl, land scarcity, and infrastructure availability, also play a crucial role in determining land values in Chennai city. [14] gave preference to infrastructure facilities such as water, electricity, transportation, and medical facilities, which impact property values in Unity Estate, Lagos Metropolis. The study ranked physical infrastructure as the most critical facility impacting property values [14]. A study examined the impact of BRT on the land market

in Seoul, Korea, and concluded that improvements to the BRT system compelled property owners to convert single-family houses into higher-density apartments [15]. Other research examined residential, office, and retail property values around the new stations from 2001 to 2008 and found that the property values were higher near the CBD area and lower in the catchment areas [16]. The regulations and zoning also impact property values, as studied by McMillen and McDonald (2004). Studies conducted by Umar Ismail (2021) and Oloke et al. emphasized security as an attribute that affects property values. The study by [17] examined the impact on the property values of Beijing's residential, commercial, and retail properties. The study observed that the transit system near the CBD had a greater influence than properties located away from the CBD. The research conducted by Bartholomew and Ewing found a significant influence of transit-oriented development on property values, examining proximity and accessibility parameters of transit stations and policies related to TOD. stations and policies related to TOD. Property values near the station area were found higher than in the area along the transit line. Residential, commercial, and industrial property values were also higher within the 1-mile radius of the station [18]. The study found that the Light Rail System positively affected property prices, but it was also associated with negative impacts, such as noise pollution and traffic congestion [19]. A study investigated key factors influencing residential property values in the Auckland, New Zealand, property market and ranked variables. The study utilized 18 independent variables and one dependent variable, and the top 5 most important attributes were the number of bedrooms, location, land size, and construction type. The availability of a garden is listed as the most significant attribute that impacts property values, whereas the size and number of bedrooms have a higher impact [3]. Urban infrastructure strongly influences property values. Also, along with infrastructure availability, quality plays a critical role in determining property values [20] [21]. Therefore, property values may depreciate due to infrastructure degradation. The various factors identified in the literature review were shortlisted into 37 critical factors, which are explained in detail below.

Table 2. List of the factors identified from the Literature

S.No.	Identified Factors
1	Location
2	Accessibility to Public Transport (Metro, Rail & Bus)
3	Accessibility to Public Amenities
4	Accessibility to the CBD
5	Accessibility to the Retail Area
6	Accessibility to Health Facilities
7	Literacy Rate
8	Population & Population Change
9	Crime Rate/Security & Safety
10	Water Scarcity
11	Pollution (Noise, Air, and Water)
12	Flood Zone
13	Urban Heat Island
14	Availability of Water Supply, Sanitation, Electricity, and Waste Disposal Systems
15	Availability of Transportation Facility (BRTS/MRTS/Airport)
16	Availability of Commercial/Market
17	Availability of Open/Green/Recreational Space
18	Availability of social infrastructure (Educational facilities -School/College/ Healthcare Facilities)
19	Current & Future Land Use
20	History of Ownership
21	Land Scarcity
22	Neighbourhood Characteristics
23	Distance from Public Transport (Airport, Bus, Metro, and Rail)
24	Distance from Solid waste, landfills, sewage treatment plants, and polluted drains
25	Distance to school/health
26	Distance from CBD
27	Distance from Park

28	Government Policies (TOD/TDR/FAR)
29	Urban Sprawl
30	Local Market Trend (Pricing & Appreciation)
31	Unsold Housing Stock/Vacancy
32	Structure of property
33	Age of the building
34	Buyer's Income
35	Open and green within the building (Structural Attribute)
37	Area of the building

Literature related to the classification of factors

The various factors that influence property values can be categorized into several groups, such as political, contextual, managerial, economic, and physical attributes [18]. Other studies identify and group factors into four categories: structural, physical, neighborhood, and environmental [19]. Some literature classified the factors into two main categories: microeconomic and macroeconomic in the case of commercial property values [20]. While other literature has clubbed the various factors into micro, meso, and macro factors based on the influence of the factors [21][22]. The factors can be classified into Intrinsic Variables and Extrinsic Variables. Intrinsic variables included the interior features of the building along with the spatial features like the number of bedrooms, number of bathrooms, construction quality, built-up area, availability of parking, etc., whereas extrinsic factors included economic growth, urbanization, Population, GDP, Employment growth, per capita income, Bank Credit, investment demand, and regional prosperity [23]. As per the study conducted, 28 attributes were identified and classified into four major categories: locational factors, Land attributes, Demographics, and neighborhood attributes[24].

Methods adopted for various studies

The identification of factors influencing property values involves a variety of research methods customized to the specific context and objectives of the study. In Delhi, India's National Capital Territory, a combination of structured and semi-structured interviews was employed to develop a Relative Importance Index (RII) and identify key factors through rank correlation analysis. Similarly, a three-cycle Delphi survey and RII method were employed in the urban areas of Kerala, engaging experts from various fields, including property valuation, real estate, and urban planning. The criteria were rated on a seven-point Likert scale until everyone agreed, with road access and proximity to disturbing sources being the most important ones [8]. In Jakarta, a structural equation model (SEM) was employed to analyze 628 property data points gathered by the Directorate General of State Asset Management, effectively recognizing important property characteristics and calculating the impact of unobserved variables on value [25]. In California, multiple linear regression was applied to examine 1,000 samples to understand multicollinearity among explanatory variables. This technique identified geographical location, housing median age, population, total bedrooms, and median income as notable factors. A Geographical Information System (GIS) was employed to create a valuation model and collect the data of spatial variables that affect the properties of urban local areas. This model featured a coefficient of correlation analysis for identification of the strongest connections, indicating that distance to the central city, distance to major roads, and width of access roads had the most notable impact [26]. These varied methodologies underscore the need for context-specific strategies to recognize and assess the factors that impact property values accurately. The research by Bender et al. 1999 employed the Analytic Hierarchy Process (AHP) following a questionnaire survey to identify the environmental quality parameters and their effect on property values in Geneva, Switzerland. A study by [27] employed a quantitative method for analyzing data collected from semi-structured interviews, identifying key factors that impact property values. A valuation model utilizing a GIS has been developed for an urban area to analyze the influence of spatial features on property values. Additionally, a study of the coefficient of correlation was conducted to determine positive aspects of relationships with the property's value [28]. The hedonic pricing model (HPM), spatial autoregressive model (SAR), and error term (SEM) were employed to quantify the study area and assess variables influencing commercial property value in a region surrounding Xi'an metro station [29].

RESULTS & ANALYSIS

Based on literature studies, 104 factors that impact residential, commercial, and other property values have been identified. With a questionnaire-based survey involving experts in a two-cycle Delphi survey, the factors are ultimately filtered down to 37. They are clubbed together under the heading of ten main attributes under the umbrella of 3 categories based on a scale of their impact, which is micro-scale, mesoscale, and macro-scale, and according to the intensity of the scale. Micro-scale parameters are categorized into structural attributes. The main attributes, including location, DCR, policy, infrastructure, real estate dynamics, connectivity, links, and legal considerations, are grouped under the mesoscale. The other attributes, which include socio-economic profile and demographics, are grouped under the macro scale, along with connectivity and linkages, as well as legal considerations. The 10 main factors include connectivity and linkage, location, infrastructure facilities, physical and social aspects, legal considerations, policy and development control regulations, real estate dynamics, structural attributes, demographics, and environmental factors.

Location

Location is the primary factor, according to the expert survey and analysis, with an average weight (AW) of 3.3, a Relative Ranking (RR) of 1, and a weight (W) of 0.161 among the 10 primary factors. The six significant sub-factors that are clubbed and analyzed under location are neighborhood characteristics, distance from public transport like Metro, bus, and Airport, distance from school/health facilities, distance from CBD, distance from Solid waste, Landfill, and sewage treatment plants & polluted drains, and distance from the park. As per analysis, it is concluded that neighborhood characteristics are the main sub-factor with, RR=1 & W=0.224, followed by distance from public transport like Metro, bus, and Airport with RR=2 & W=0.212, Distance from Solid waste, Landfill, and sewage treatment plants & polluted drains (RR=3 & W=0.196), distance from school/health facilities (RR=4 & W=0.184), distance from CBD (RR=5 & W=0.183), distance from health facilities & educational facilities (RR=5 & W=0.183) and distance from the park (RR=6 & W=0.168). Location is a factor influenced by various sub-factors, as shown in Equation. 1.

$$F1\,(Location) = 0.224x(L1) + 0.212x(L2) + 0.196x(L3) + 0.184x(L4) + 0.183x(L5) + 0.168x(L6) \quad (1)$$

Connectivity & linkages

Under the category of connectivity and linkage, five further sub-attributes are listed: accessibility to public transport, accessibility to health facilities, accessibility to the CBD, accessibility to retail areas, and accessibility to public utilities. The relationship of the factor with its sub-factors is shown in Equation 2.

$$F2\,(Connectivity\ \&Linkages) = 0.229x(CL1) + 0.207X(CL2) + 0.195X(CL3) + 0.186x(CL4) + 0.181x(CL5) \quad (2)$$

According to the analysis, connectivity and linkage are the second most important factors, as the average weight assigned by the expert is 3.2, the ranking score is 2, and the weight is 0.151. There are five sub-factors, which are categorized under the category accessibility to public transport (metro, railway, and bus) has the first rank and weight 0.229, accessibility to public amenities (RR=2 & W=0.207), accessibility to CBD (RR=3 & W=0.195), accessibility to health facilities (RR=4 & W=0.186), and accessibility to Schools/Colleges/ Universities (RR-5 & W=0.181).

Infrastructure Facilities

Infrastructural facilities (physical and social) are one of the factors that scored RR=3 & W=0.122. This category consists of water supply, sanitation, electricity, and waste disposal with RR= 1& W=0.178, transportation facilities with RR=2 & W=0.178, health facilities with RR=5 & W=0.155, commercial/market (RR=4 & W=0.141), school/ colleges (RR=5 & W=0.145), and open/ recreational spaces (RR=3 & W=0.175), which is shown in Equation 3.

$$F3(Infrastructural\ Facilities) = 0.178x(IF1) + 0.178x(IF2) + 0.175x(IF3) + 0.167x(IF4) + 0.145x(IF5) \quad (3)$$

Structural Attributes

Structural attributes are also primary attributes for a buyer. Design, area, age, and structure of the property are the sub-factors that fall under this category. According to the analysis, structural attributes are scored as RR=4 and W=0.116. Among the sub-factors, the structure of the building is ranked first with W=0.297, followed by the area of the building (RR=2, W=0.380), the age of the property (RR=3, W=0.379), and design (RR=4, W=0.250). Therefore, the structural attribute is a combination of the four sub-attributes, and the relationship between the factors is illustrated in the Equation 4.

$$F4(Structural\ Attributes) = 0.397x(SA1) + 0.380Xx(SA2) + 0.379x(SA3) + 0.250x(SA4) \quad (4)$$

Policy & Development Control Regulations (DCR)

Policy and development control regulation (DCR) is one of the parameters with RR=5 and W=0.099807. Three sub-factors are grouped under this category: government policies (RR=1 & W=0.335), land use and zoning (RR=2 & W=0334), and urban sprawl (RR=3 & W=0.329), and the relationship between the attributes and their sub-attributes is shown in Equation 5.

$$F5\ (Policy\ \&DCR) = 0.335x(PD1) + 0.334x(PD2) + 0.329x(PD3) \quad (5)$$

Real Estate Dynamics

Real estate dynamics is at rank 6 with a weight of 0.091. Two sub-factors are grouped under this factor, which are a local market trend (pricing & appreciation) and unsold housing stock (RR=1, W=0.517), and (RR=2, W=0.482), respectively, as shown in Equation 6.

$$F6\ (Real\ Estate\ Dynamics) = 0.517x(RED1) + 0.482x(RED2) \quad (6)$$

Legal Considerations

This category of Legal consideration consists of three sub-attributes: current and future land use (RR = 1, W = 0.347), history of ownership (RR = 3, W = 0.338), and land scarcity (RR = 4, W = 0.314). Out of three sub-attributes, current and future land use is the most significant, and land scarcity is the least significant attribute. The relationship between the factor and sub-factors is shown in Equation. 7.

$$F7(Legal\ Consideration) = 0.347x(LC1) + 0.338x(LC2) + 0.314x(LC3) \quad (7)$$

Demographics

Demography ranks 8th with a weight of 0.071. The sub-attributes under demographics are population and population change, literacy rate, and income category. According to the expert survey and analysis, Income (RR = 1 and W = 0.351) is the primary sub-factor, followed by literacy rate (RR = 2 and W = 0.339) and population and population change (RR = 3 and W = 0.310). The relationship is as shown in the Equation. 8.

$$F8(Demographics) = 0.351x(D1) + 0.339x(D2) + 0.310x(D3) \quad (8)$$

Socio-Economic Profile

The socioeconomic profile, as an attribute, has a relative ranking of 9 and an overall weight of 0.059. Three sub-factors have their scores as crime rate and security (RR = 1 and W = 0.371), buyer's income (RR = 2 and W = 0.371), and employment (RR = 3 and W = 0.267). The relationship of the subfactors with the main factor is shown in the Equation. 9.

$$F9(Socio-Economic\ Profile) = 0.371x(SEP1) + 0.371x(SEP2) + 0.267x(SEP3) \quad (9)$$

Environmental Factor

The attribute comprises four sub-attributes that negatively affect the property's value. The environmental factor is ranked 10th with a weight of 0.0579. Water Scarcity (RR=1 & W=0.262), Pollution levels (RR=2 & W=0.258), flood zone (RR=3 & W=0.255), and heating (RR=4 & W=0.245) are four subfactors under the Environmental factor, with UHI having the least impact. The relationship is shown in Equation. 10.

$$F10(Environmental\ Factors) = 0.262x(EF1) + 0.258x(EF2) + 0.254x(EF3) + 0.255x(EF4) \quad (10)$$

The Property values are the aggregation of the factors and sub-factors that influence the intensity of the main factor's impact, as shown in the Equation. 11.

$$PV = 0.161(F1) + 0.151(F2) + 0.122(F3) + 0.116(F4) + 0.099(F5) + 0.091(F6) + 0.072(F7) + 0.071(F8) + 0.059(F9) + 0.58(F10) \quad (11)$$

Where PV-Property Value, F1-Location, F2-Connectivity & Linkages, F3-Infrastrcuture Facilities (Physcial &Social), F4- Structural Attributes, F5-Policy and Developmemt Control Regulation, F6- Real Estate Dynamics, F7-Legal Considetation, F8- Demographics, F9- Socio-Economics Profile and F10- Environmental Factors

According to the analysis, it is concluded that property value is a function of location attributes, connectivity and linkages, availability of infrastructural facilities, structural attributes, policy and development control regulations, real estate dynamics, legal considerations, demographics, socio-economics, and environmental factors, as shown in Fig 2. Among all the factors, location, connectivity & linkages are the primary factors that impact property values overall. Neighborhood characteristics (4.67) have the highest weight within the sub-factors of location, followed by distance from public transport (4.47).

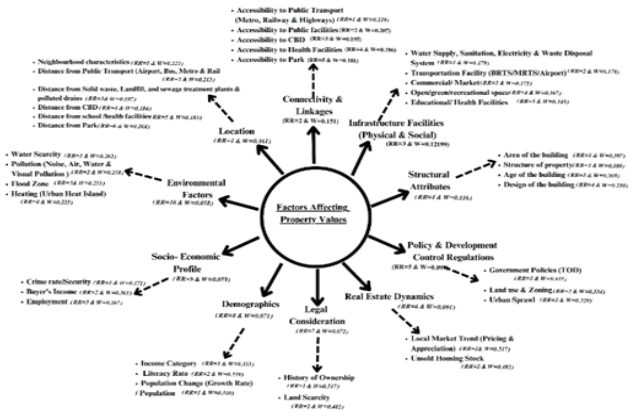


Fig 2: List of Factors and sub-factors affecting property values with their weights

Distance from a park has the lowest weight, at 3.85. Connectivity and linkage have a weight of 3.2 and a relative ranking of 2 out of 10 main attributes. Availability of public transport has the highest relative ranking out of 5 sub-factors of Connectivity & Linkages, and accessibility to parks has the lowest, i.e., 5. Infrastructure Facilities (Physical and Social) received a weight of 2.5 and a relative rank of 3, followed by structural attributes, Real Estate dynamics, demographics, and socioeconomic profile. Environmental factors have the lowest weight and ranking. The results show that property value doesn't depend on one factor but on the cumulative effect of various parameters. The proposed equation provides a comprehensive decision-making framework that can be used by policymakers, real estate consultants, developers, and various financial institutions for the assessment of any project.

CONCLUSION

The study clearly proves that the location and connectivity & linkage are the primary factors that impact the property value, which proves the classical urban economic theories like the Bid Rent Theory. It was discovered that the Delphi technique was effective in determining and examining the critical factors that affect the values. The relative ranking method was utilized to identify the relative significance of variables. All identified factors impact property values to some extent. Properties in desirable locations tend to have higher values because of their proximity to amenities, schools, shopping centers, and other conveniences. Additionally, properties in areas with better connectivity, easy access to public transportation, good job opportunities, and low crime rates often command higher prices. Regarding the impact on property value, MRTS has a greater effect than other available infrastructure, such as BRTS. Metro presence has increased the achievable FSI/FAR, resulting in higher real estate values. DCR (Development Control Regulation) is very important for the applicability of TOD. Therefore, Government policies play a vital role in the infrastructure, which directly or indirectly impacts property values. Structural attributes of a property also affect its value. The size, layout, number of bedrooms and bathrooms, and the property's overall condition can impact its market value. Despite its exploratory character, this study offers insight into the identification and selection

of parameters that determine property value. The study's methodology is believed to aid researchers in adopting comparable approaches in various Indian and international marketplaces. Additionally, it is expected that the investigation outcomes will help working professionals in various fields such as real estate development, property management, urban planning, banking, and local governments identify the crucial influencing factors while making decisions. Only the most significant factors have been identified in this research; can directly be included in further analysis, which can reduce the workload associated with gathering a large amount of data. Also, this variable selection strategy may be replicated in similar circumstances and may be useful when data is limited. To create property value models, further study is needed to statistically evaluate the identified factors and determine the relative importance of each component on property value.

ACKNOWLEGEMENT

The authors are thankful to the departmental research committee members for their regular reviews and suggestions, and the parent organization for allowing us to conduct the research in this domain.

REFERENCE

1. P. Taltavull and D. La Paz, "Interest rates and Housing Cycle : the case of Southern Europe," 2002.
2. D. K. Hayunga and R. K. Pace, "Spatial statistics applied to commercial real estate," J. Real Estate Financ. Econ., vol. 41, no. 2, pp. 103–125, 2010, doi: 10.1007/s11146-009-9190-2.
3. X. J. Ge and Y. Du, "Main variables influencing residential property values using the entropy method – The case of Auckland," 5th Int. Struct. Eng. Constr. Conf., pp. 2–21, 2007, [Online]. Available: http://www.asres2007.umac.mo/papers/041 - PAPER.pdf
4. G. Malhotra and A. Rastogi, "Impact of Metro Rail Transit System on Residential Properties," Internatioanl J. Creat. Reserach Thoughts, vol. 8, no. 5, pp. 1534–1540, 2020.
5. A. E. Nwosu, "Assessement of the effect of urban infrastrcuture as a tool for enhancing the values of residential property values," Ethiop. J. Environ. Stud. Manag., vol. 8, no. 2, pp. 196–205, 2016, doi: doi: http://dx.doi.org/10.4314/ejesm.v8i2.9.
6. C. Mulley and C. H. (Patrick) P. Tsai, "When and how much does new transport infrastructure add to property values ? Evidence from the bus rapid transit system in Sydney , Australia," Transp. Policy, vol. 51, pp. 1–9, Oct. 2016, doi: 10.1016/j.tranpol.2016.01.011.
7. M. Singh and R. Sharma, "Financing Options for a Transit System through Real Estate," in European Transport Conference 2012, 2012.
8. B. BV, N. MA, and A. K. PP, "A methodology for identifying critical factors influencing land value in urban areas: a case study of Kerala, India," Prop. Manag., vol. 38, no. 5, pp. 665–681, 2020, doi: 10.1108/PM-01-2020-0004.
9. N. J. Shariff, "Utilizing the Delphi Survey Approach: A Review," J. Nurs. Care, vol. 04, no. 03, pp. 246–251, 2015, doi: 10.4172/2167-1168.1000246.
10. S. Hanafin, "Review of literature on the Delphi Technique," Dublin Natl. Child. Off., no. March, pp. 1–51, 2004, [Online]. Available: http://www.childrensdatabase.ie/documents/publications/Delphi_Technique_A_Literature_Review.pdf
11. G. Malhotra and A. Rastogi, "Impact of Metro Rail Transit System," vol. 8, no. 5, pp. 1534–1540, 2020.
12. M. Zhang and L. Wang, "The impacts of mass transit on land development in China: The case of Beijing," Res. Transp. Econ., 2013, doi: 10.1016/j.retrec.2012.06.039.
13. S. Swamidurai, "Factors Affecting Urban Land Value in Indian Cities - Chennai City as a Case Study," Int. J. Res. Sci. Innov., vol. 1, no. 3, pp. 31–34, 2014.
14. A. M. O and A. O. O, "Effects of Infrastructure on Property Values in Unity Estate, Lagos, Nigeria," Int. J. Econ. Manag. Soc. Sci., vol. 2, no. 5, pp. 195–201, 2013.
15. F. Pagliara and E. Papa, "Urban rail systems investments : an analysis of the impacts on property values and residents ' location," J. Transp. Geogr., vol. 19, no. 2, pp. 200–211, Mar. 2011, doi: 10.1016/j.jtrangeo.2010.02.006.
16. Y. Gu, "The impacts of rail transit on property values and land development intensity: The case of No.13 Line in Beijing," in AsRes Conference. 2007, 2007.
17. K. and X. (Jason) Cao, "Transitway Impacts Research Program The Hiawatha Line : Impacts on Land Use and Residential Housing Value," 2010.

18. C. Atkinson-Palombo, "Comparing the capitalisation benefits of light-rail transit and overlay zoning for single-family houses and condos by neighbourhood type in metropolitan Phoenix, Arizona," Urban Stud., vol. 47, no. 11, pp. 2409–2426, 2010, doi: 10.1177/0042098009357963.

19. A. A. Adebayo, P. Greenhalgh, and K. Muldoon-Smith, "Investigating retail property market dynamics through spatial accessibility measures," J. Eur. Real Estate Res., vol. 12, no. 2, pp. 155–172, 2019, doi: 10.1108/JERER-01-2018-0009.

20. E. K. Chikwuado, "Alternative Measure in Providing Real Estate Infrastructure in a Recess Economy in Awka, Anambra State, Nigeria," Int. J. Eng. Appl. Sci. Technol., vol. 04, no. 11, pp. 28–33, 2020, doi: 10.33564/ijeast.2020.v04i11.006.

21. and G. P. Avellaneda, Claudia N., "Determinants of property value reappraisals: municipal responsiveness to urban changes," in What Works in Latin American Municipalities?, C. Avellaneda, Ed., 2023, ch. 6, pp. 149–201. doi: https://doi.org/10.4337/9781803929071.00014.

22. O. Ruza, O. Lavrinenko, and R. Zelcs, "Sustainable Development of real estate market : Impact of the Micro and Meso Level Factors," J. Secur. Sustain. Issues, vol. 3, no. 4, pp. 45–60, 2014, doi: http://dx.doi.org/10.9770/jssi.2014.3.4(5).

23. N. S. Razali, N. I. Menan, S. V. Teng, N. L. Ting, N. H. Mahayuddin, and M. S. A. Rahman, "Factors affecting industrial property value," Int. J. Sci. Technol. Res., vol. 9, no. 1, pp. 212–217, 2020.

24. S. S. Sandbhor and N. B. Chaphalkar, "State of Art Report on Variables Affecting Housing Value," Indian J. Sci. Technol., vol. 9, no. April, 2016, doi: 0.17485/ijst/2016/v9i17/82861.

25. Y. Gu, "The Impacts of Rail Transit on Property Values: Empirical Study in Beijing," SSRN Electron. J., no. 60, pp. 1–15, 2011, doi: 10.2139/ssrn.1108982.

26. K. A. Prasetyo, D. A. Safitra, and A. P. Swasito, "Identification of factors influencing land value for state's assets mass appraisal purposes: Evidence from Indonesia," Plan. Malaysia, vol. 19, no. 3, pp. 37–47, 2021, doi: 10.21837/PM.V19I17.985.

27. T. Li and X. Yang, "The research on factors influencing house value-take California as an example," Theor. Nat. Sci., vol. 39, no. 1, pp. 96–102, 2024, doi: 10.54254/2753-8818/39/20240571.

28. S. Senadheera and E. Warusavitharana, "A Property valuation model to identify thriving real estate opportunities, based on spatial factors," in Independence and Interdependence of Sustainable Spaces, 2022, pp. 186–195. doi: 10.31705/faru.2022.21.

29. S. Singhal and Y. Tyagi, "Analyzing the Influence of Metro Stations on Commercial Property Values in Delhi: A Hedonic Approach," Real Estate Manag. Valuat., vol. 29, no. 4, pp. 10–22, 2021, doi: 10.2478/remav-2021-0026.

DOI - 10.68648/ijte-q-2026-49-03-002

Optimisation of Biogas Production from Co-Digestion of Organic Waste Streams of Kolhapur Municipal Corporation Using Response Surface Methodology

Prakash Lagamavar, Sudhir B. Desai, Mrunalini P. Jagtap
Yashwantrao Chavan School of Rural Development
Shivaji University, Kolhapur, Maharashtra
✉ prakash@srigrowgas.com
✉ desaisudhirb@gmail.com
✉ powar.mrunalini4@gmail.com

ABSTRACT

Kolhapur Municipal Corporation (KMC) generates approximately 488 tonnes of municipal solid waste daily, of which nearly 57% is biodegradable organic matter presently disposed of without energy recovery. This study characterises five major organic waste streams of KMC — food waste, vegetable market waste, fruit market waste, cattle dung, and slaughterhouse waste — and optimises their co-digestion ratio using Response Surface Methodology (RSM) with a D-optimal mixture design. Individual substrate characterisation revealed complementary carbon-to-nitrogen (C/N) profiles, motivating a co-digestion strategy. A 38-run mixture design was executed in bench-scale CSTR digesters, yielding a statistically significant model ($F = 48.52$, $R^2 = 0.981$). The optimised blend — 40% food waste, 25% vegetable waste, 20% fruit waste, 10% cattle dung and 5% slaughterhouse waste — achieved a C/N ratio of 24.8:1 and a specific biogas yield of 724 mL/g VS, validated experimentally within 1% of the model prediction. A conceptual 200 MT/day plant based on this formulation shows techno-economic viability (IRR 18.4%, payback period 6.8 years). The findings demonstrate that systematic co-digestion optimisation, grounded in a rigorous waste characterisation and experimental design methodology, can substantially improve biogas recovery from municipal organic waste in Tier-2 Indian cities.

KEYWORDS *: Anaerobic co-digestion; Biogas; Municipal solid waste; Response Surface Methodology; Kolhapur*

NOMENCLATURE

AD	Anaerobic Digestion
BMP	Biochemical Methane Potential
CD	Cattle Dung
COD	Chemical Oxygen Demand
CSTR	Continuously Stirred Tank Reactor
C/N	Carbon-to-Nitrogen Ratio
FAN	Free Ammonia Nitrogen
FCO	Fertiliser Control Order
FOS/TAC	Volatile Acids to Alkalinity Ratio
FrMW	Fruit Market Waste
FW	Food Waste
HRT	Hydraulic Retention Time
IRR	Internal Rate of Return
ISR	Inoculum-to-Substrate Ratio
KMC	Kolhapur Municipal Corporation
MSW	Municipal Solid Waste
NPV	Net Present Value
OLR	Organic Loading Rate
RSM	Response Surface Methodology
SHW	Slaughterhouse Waste
TS	Total Solids
VFA	Volatile Fatty Acids
VMW	Vegetable Market Waste
VS	Volatile Solids

INTRODUCTION

Municipal solid waste (MSW) generation has grown steadily worldwide with urbanisation, and Indian cities are witnessing a particularly sharp rise, with national MSW generation projected to increase substantially over the coming decade as urban population expands [13]. In response, the Government of India has promoted several waste-to-energy initiatives, including the GOBARdhan (Galvanising Organic Bio-Agro Resources Dhan) scheme and the SATAT (Sustainable Alternative Towards Affordable Transportation) initiative, both aimed at converting organic waste and cattle dung into compressed biogas (CBG) and organic manure under the National Bioenergy Programme [14, 25]. Despite this policy push, conversion infrastructure at the municipal level remains limited in most Tier-2 cities, and a large share of biodegradable waste continues to be transported to open dumping grounds without energy recovery.

Kolhapur Municipal Corporation (KMC), a Class-A civic body serving approximately 6.5 lakh residents across 76 administrative wards spanning 67.77 km², generates about 488 MT of solid waste per day, of which nearly 57% (roughly 278 MT/day) is biodegradable organic matter [27]. At present this fraction is collected and transported to the Kasaba Bawada disposal site with limited energy recovery, representing a significant lost opportunity for renewable energy generation and organic fertiliser production.

Anaerobic digestion (AD) is a biologically mediated, four-stage process — hydrolysis, acidogenesis, acetogenesis and methanogenesis — through which complex organic polymers are sequentially broken down by consortia of bacteria and archaea into biogas, typically comprising 50–70% methane (CH_4) and 30–45% carbon dioxide (CO_2) with trace hydrogen sulphide (H_2S) and moisture [22]. The residual digestate, rich in nitrogen, phosphorus and potassium, can be mechanically separated into solid and liquid fractions and applied as organic fertiliser, closing the nutrient loop between urban waste generation and peri-urban agriculture [23].

Mono-digestion of a single waste stream is, however, frequently unstable because individual streams rarely possess a carbon-to-nitrogen (C/N) ratio in the optimal range of 20–30:1 required for balanced microbial activity: nitrogen-rich streams risk ammonia inhibition, while carbon-rich streams digest slowly and yield poorly [8, 17]. Co-digestion of complementary waste streams is widely reported to overcome this limitation by diluting inhibitory compounds, balancing nutrients and improving buffering capacity [4].

The objectives of this study are to: (i) characterise the physico-chemical properties of the five principal organic waste streams generated within KMC; (ii) determine their individual biogas yield potential through biochemical methane potential (BMP) assays; (iii) screen two-substrate combinations to identify synergistic pairings; (iv) statistically optimise a five-substrate co-digestion recipe using Response Surface Methodology (RSM) to maximise specific biogas yield; and (v) evaluate the conceptual plant design and techno-economic feasibility of a full-scale facility based on the optimised formulation.

LITERATURE REVIEW

Global Studies on Anaerobic Co-Digestion of Municipal Organic Waste

Elango et al. [1] investigated biogas production from MSW combined with domestic sewage in a batch digester operated at 26–36 °C with a fixed hydraulic retention time (HRT) of 25 days, and reported a maximum yield of 0.36 m³/kg VS at an optimum organic feeding rate of 2.9 kg VS/m³/day, alongside 87.6% total solids reduction. Mata-Alvarez et al. [4], in a meta-analysis of over 1,200 co-digestion studies published between 2010 and 2013, confirmed that blending complementary substrates generally improves methane yield relative to mono-digestion through nutrient balance and dilution of inhibitory compounds. Bouallagui et al. [5] found that combining vegetable market waste with cattle dung at a 3:1 ratio improved the blended C/N ratio to 22.5:1 and raised biogas yield to 0.41 m³/kg VS, while Wan Azlina and Idris [2] demonstrated that leachate chemical oxygen demand (COD) strength strongly influences biogas production kinetics in MSW leachate co-digestion. Pavi et al. [7] demonstrated that co-digesting the organic fraction of MSW with fruit and vegetable waste at an optimal ratio of 1:3 increased methane yield by 141% relative to mono-digestion of MSW alone, achieving 396.6 mL

CH_4/g VS — consistent with the broader finding that food-waste-rich blends are among the most readily biodegradable co-digestion feedstocks [24]. Pognani et al. [10] and Scaglia et al. [11] separately reported that substituting energy crops with the organic fraction of MSW in full-scale digesters did not compromise specific biogas yield, while Abad et al. [12] showed that characterising the biogas potential of multiple industrial waste streams enables cost-optimised co-digestion recipes in a circular-economy framework.

Co-Digestion Studies and Policy Context in India

Singh et al. [6], working with waste from the Azadpur mandi in Delhi — Asia's largest wholesale fruit and vegetable market — reported that a 2:1 vegetable-waste-to-cattle-dung ratio improved specific yield to 0.58 m^3/kg VS. Rajendran et al. [15] reviewed household-scale biogas digesters across developing countries and highlighted that feedstock variability, rather than digester design, is frequently the dominant constraint on performance in decentralised Indian systems. Kondusamy and Kalamdhad [16] reviewed pre-treatment strategies for food waste digestion and noted that Indian food waste, owing to its high oil and salt content relative to Western counterparts, often requires blending with a buffering co-substrate such as cattle dung or agricultural residue to sustain stable methanogenesis. At the policy level, the Ministry of New and Renewable Energy's National Bioenergy Programme and the GOBARdhan scheme [14, 25] provide capital subsidy and viability-gap-funding mechanisms for municipal biomethanation and CBG projects, but implementation has been constrained in Tier-2 cities by the absence of locally validated feedstock-blending protocols — a gap this study addresses directly for Kolhapur.

Process Inhibition, C/N Ratio and Ammonia Toxicity

Yeniğün and Demirel [8] established that free ammonia nitrogen (FAN) concentrations above 100–200 mg/L are strongly inhibitory to methanogenic activity, a finding directly relevant to managing nitrogen-rich streams such as slaughterhouse waste. Chen et al. [17], in a widely cited review of anaerobic digestion inhibition mechanisms, catalogued ammonia, volatile fatty acid (VFA) accumulation, sulphide, and light metal ions as the principal inhibitors of methanogens, and emphasised that co-digestion is among the most effective strategies for diluting any single inhibitor below its threshold concentration. Khalid et al. [26] similarly concluded, in a review of solid-waste anaerobic digestion, that blending nitrogen-rich and carbon-rich substrates to approach a C/N ratio of 20–30:1 is the single most consistent predictor of stable, high-yield digestion across the studies they surveyed.

Statistical Optimisation of Anaerobic Digestion Using RSM

Response Surface Methodology is a collection of statistical and mathematical techniques for developing, improving and optimising processes in which a response of interest is influenced by several variables, and is particularly suited to mixture problems where components must sum to a fixed total [21]. Karthik Rajendran et al. [3] applied a techno-economic simulation framework to biogas production from the organic fraction of MSW and demonstrated that systematic scenario optimisation, rather than fixed design heuristics, materially improves both yield and project economics. While RSM and related design-of-experiments approaches are well established in bioprocess and wastewater engineering, their application to multi-substrate co-digestion of Indian municipal waste streams specifically remains limited in the published literature.

Research Gap

Despite this substantial literature base, no published study has systematically characterised the full set of organic waste streams generated in a Tier-2 Indian city and optimised their co-digestion ratio using a rigorous statistical mixture-design framework validated through both batch and semi-continuous trials, and further extended into a conceptual plant design with techno-economic assessment. This study addresses that gap for Kolhapur Municipal Corporation, and proposes a methodology that is directly transferable to comparable Indian cities.

METHODOLOGY

Study Area and Waste Sampling Protocol

Five organic waste streams generated within KMC were selected for study on the basis of their contribution to total organic waste tonnage: food waste (FW) from hotels, messes and households; vegetable market waste

(VMW) from the Mahadwar Road, Rajarampuri and Shahupuri market clusters; fruit market waste (FrMW) from the wholesale fruit market and five associated retail clusters; cattle dung (CD) from four dairy cooperatives and two goshalas within KMC limits; and slaughterhouse waste (SHW) from the KMC-operated municipal slaughterhouse. Representative composite samples of each stream (approximately 5 kg per sample) were collected weekly over a twelve-week period spanning two seasons — monsoon (July–September) and winter (January–March) — to capture seasonal variability in waste composition. Samples were manually sorted to remove inert and non-biodegradable contraries, size-reduced to a nominal particle size of below 10 mm using a laboratory shredder, homogenised, and either analysed within 48 hours of collection or stored at 4 °C to minimise pre-analysis biodegradation.

Analytical Methods and Standard Protocols

Each waste stream was characterised for total solids (TS), volatile solids (VS), C/N ratio, pH and chemical oxygen demand (COD) in triplicate, using the standard methods summarised in Table 1. TS and VS were determined gravimetrically by oven-drying at 105 °C to constant weight followed by ashing at 550 °C, per APHA Standard Methods [20]. Total organic carbon and total Kjeldahl nitrogen were determined using a CHNS elemental analyser and the Kjeldahl digestion-distillation method respectively, from which the C/N ratio was calculated. pH was measured using a calibrated bench-top pH meter on a 1:10 (w/v) aqueous extract, and COD was determined by the closed reflux titrimetric method.

Table 1: Analytical parameters and standard methods used for waste characterisation

Parameter	Method / Standard	Instrument
TS, VS	Gravimetric (105 °C / 550 °C) – APHA 2540 B, E	Hot-air oven, muffle furnace
C/N ratio	CHNS analysis + Kjeldahl digestion	CHNS elemental analyser
pH	Electrometric, 1:10 (w/v) extract	Digital pH meter
COD	Closed reflux titrimetric – APHA 5220 C	COD digester, titration
VFA, Alkalinity	Titrimetric (Nordmann method)	Auto-titrator
Biogas composition	Gas chromatography (TCD/FID)	Portable biogas analyser / GC

Biochemical Methane Potential (BMP) Assay

The biogas yield potential of each individual substrate was determined using a batch BMP assay adapted from the protocol of Angelidaki et al. [18] and VDI 4630 [19]. Anaerobic sludge from an operating digester, acclimatised for two weeks prior to use, served as inoculum. Batch reactors (500 mL serum bottles, 400 mL working volume) were charged with substrate and inoculum at an inoculum-to-substrate ratio (ISR) of 2:1 on a VS basis, flushed with nitrogen gas to establish anaerobic headspace conditions, sealed, and incubated at 35 ± 1 °C in a temperature-controlled shaking incubator for up to 30 days, or until daily gas production fell below 1% of the cumulative volume. Biogas volume was measured daily by water displacement, and composition (%CH_4, %CO_2) was determined every three days by gas chromatography. Inoculum-only blanks were run in parallel and subtracted from substrate assay values to correct for endogenous gas production. All assays were conducted in triplicate, and results are expressed as mL CH_4/g VS added.

Screening of Two-Substrate Combinations

Prior to the full mixture-design experiment, all ten possible two-substrate combinations of the five waste streams were screened at two mixing ratios each in bench-scale 5 L borosilicate CSTR digesters (3 L working volume, 35 ± 1 °C mesophilic, low-speed continuous mixing), with biogas volume measured by water displacement over a 30-day batch cycle. A synergy index, defined as the ratio of the observed biogas yield of the blend to the VS-weighted theoretical yield of its components in isolation, was used to rank combinations and shortlist the most promising pairings for inclusion in the multi-substrate optimisation.

Experimental Design for Co-Digestion Optimisation (RSM)

A D-optimal mixture design was constructed in Design-Expert® v13 (Stat-Ease Inc.) to optimise the proportion of each of the five substrates, expressed as a fraction of

total VS subject to a sum-to-100% mixture constraint: FW (X_1: 20–60%), VMW (X_2: 10–35%), FrMW (X_3: 5–30%), CD (X_4: 5–25%), and SHW (X_5: 2–15%). The design comprised 38 experimental runs, including model points, five centre-point replicates for pure-error estimation, and additional lack-of-fit points, generated to support fitting of a special-cubic mixture polynomial of the general form:

$$Y = \Sigma\beta_i X_i + \Sigma\Sigma\beta_{ij} X_i X_j + \Sigma\Sigma\Sigma\beta_{ijk} X_i X_j X_k$$

where Y is the predicted response (specific biogas yield or methane content), X_i are the substrate proportions, and β terms are the linear, quadratic and special-cubic mixture coefficients estimated by least-squares regression. Each of the 38 runs was conducted in triplicate in the bench-scale CSTR digesters described in Section 3.4, with total biogas yield and methane content recorded as the two response variables. Model adequacy was assessed using analysis of variance (ANOVA), the lack-of-fit test, the coefficient of determination (R^2), adjusted R^2, predicted R^2, and the adequate precision ratio, which is recommended to exceed 4 for a model to be considered suitable for navigating the design space. The optimum blend was identified using a numerical desirability function (Derringer and Suich approach), which simultaneously maximises specific biogas yield and methane content subject to the mixture constraint, and the model-predicted optimum was subsequently validated by independent triplicate experimental runs at the identified blend ratio.

Semi-Continuous Digester Operation and Process Monitoring

Following identification of the optimum blend, a bench-scale CSTR (5 L total, 3 L working volume) was operated in semi-continuous (fed-batch) mode for 30 days at the design organic loading rate (OLR) and hydraulic retention time (HRT) derived from separate single-factor trials (Section 4.5); OLR and HRT are widely recognised as the two operating parameters with the greatest influence on digester stability and specific yield [9]. The digester was fed once daily, with an equal volume of digestate withdrawn prior to each feeding to maintain constant working volume. Process performance was monitored through daily biogas volume measurement, gas composition analysis by gas chromatography every three days, daily pH measurement, and the ratio of volatile organic acids to total inorganic carbonate (FOS/TAC), determined twice weekly by the Nordmann titration method. Free ammonia nitrogen was calculated from total ammoniacal nitrogen, pH and temperature using the Anthonisen equation. The digester was considered to have reached steady state when daily biogas yield varied by less than 10% over three consecutive HRTs.

Digestate Quality Assessment

Digestate from the semi-continuous trial was mechanically separated into solid and liquid fractions using a screw-press separator. The solid fraction was analysed for total Kjeldahl nitrogen (Kjeldahl method), P_2O_5 (vanadomolybdate spectrophotometric method), K_2O (flame photometry), moisture content, pH, and faecal coliform count (most-probable-number method), and assessed against the Fertiliser Control Order (FCO) 1985, Schedule VI specifications for organic fertilisers [23].

Conceptual Plant Design and Mass Balance

A conceptual biogas plant was sized on a design basis of 200 MT/day of organic waste input, corresponding to approximately 72–80% of KMC's estimated recoverable organic waste after accounting for collection efficiency and seasonal variability. Digester working volume was computed from the design OLR and the daily VS loading implied by 200 MT/day of the optimised blend, and distributed across four CSTR digesters of nominal 2,250 m^3 each, sized to standard fabrication modules and configured to allow one unit to be taken offline for maintenance without halting plant operation. Gross biogas production was extrapolated from the bench-scale specific yield with a conservative scale-up de-rating factor applied. Two downstream gas utilisation pathways were considered: combined heat and power (CHP) generation for electricity export, or upgrading to compressed biogas (CBG) via water-scrubbing or pressure-swing adsorption; the plant layout was conceived to accommodate either. Digestate management comprises windrow composting of the solid fraction (30–45 days) to produce a stabilised biofertiliser, with the liquid fraction partly recirculated as process dilution water and the remainder distributed to farmers.

Techno-Economic Analysis Methodology

Capital cost (CAPEX) was estimated using a bottom-up approach comprising civil works (digesters, tanks, foundations), mechanical equipment (mixers, pumps, gas holders, and the CHP or upgrading skid), electrical and instrumentation systems, and land and balance-of-plant infrastructure, benchmarked against typical per-tonne installed-capacity costs reported for Indian biomethanation projects. Operating cost (OPEX) components included manpower, purchased power, maintenance, chemicals and consumables, and digestate handling. Revenue streams considered were power or CBG sales, biofertiliser sales, and avoided tipping/disposal cost. Financial performance was evaluated using discounted cash flow (DCF) analysis over a 25-year project life at a 12% discount rate, representative of the cost of capital for Indian municipal infrastructure projects, from which net present value (NPV), internal rate of return (IRR) and simple payback period were derived. A one-at-a-time sensitivity analysis was performed on capital cost, feedstock/tipping cost, and power or CBG selling price to test the robustness of the computed IRR to input uncertainty.

4. RESULTS AND DISCUSSION

Waste Stream Characterisation

Individual substrate characterisation (Table 2) showed FW and VMW to be nitrogen-rich, with C/N ratios of 13.7 and 15.3 respectively; SHW was extremely nitrogen-rich (C/N = 5.9); while FrMW and CD were closer to the optimal 20–30:1 range (C/N = 26.2 and 27.0 respectively). This spread of C/N ratios across the five streams is precisely the condition under which co-digestion is expected to outperform mono-digestion of any single stream [4, 26].

Table 2: Physico-chemical characteristics and specific BMP of individual waste streams

Waste Stream	VS (%)	C/N Ratio	BMP (mL CH_4/g VS)
Food Waste	24.3	13.7	612
Vegetable Mkt. Waste	14.1	15.3	480
Fruit Mkt. Waste	17.2	26.2	525
Cattle Dung	16.7	27.0	310
Slaughterhouse Waste	29.5	5.9	485

Individual Substrate BMP and Two-Substrate Screening

BMP assays (Figure 1) ranked FW highest (612 mL CH_4/g VS), followed by FrMW (525) and SHW (485), with CD lowest (310 mL CH_4/g VS) owing to its already rumen-digested, recalcitrant fibre content — though CD remained valuable for its buffering capacity rather than its yield contribution. Screening trials identified FW combined with CD (50:50) as the strongest two-substrate synergy, improving yield by 14.8% over the weighted average of the components, consistent with the buffering mechanism described by Chen et al. [17]. Combinations exceeding 15% SHW consistently exhibited ammonia accumulation during screening, consistent with the inhibition threshold reported by Yeniğün and Demirel [8].

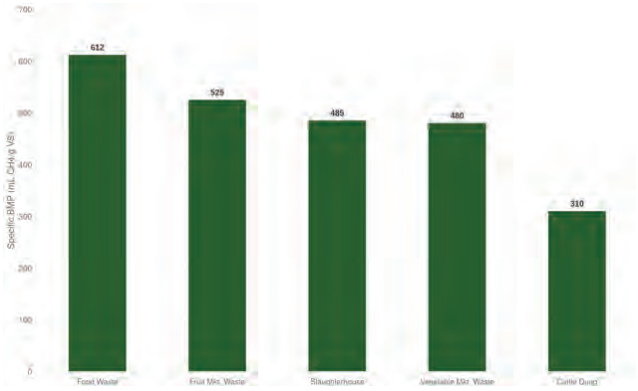


Fig. 1: Specific biochemical methane potential (BMP) of individual waste streams

RSM Model Fitting and Statistical Validation

The mixture-design model fitted to the 38 experimental runs was highly significant ($F = 48.52$, $p < 0.0001$) with $R^2 = 0.981$ and a non-significant lack-of-fit ($p = 0.184$), indicating excellent agreement between predicted and observed responses across the design space; the adequate precision ratio exceeded the recommended threshold of 4, confirming the model's suitability for navigating the mixture space. Biogas yield across the 38 runs ranged from 508–724 mL/g VS. Ranked by standardised coefficient, FrMW (+4.82) and FW (+3.95) contributed most positively to yield, VMW (+2.18) and CD (+0.86) contributed moderately, while SHW (−1.24)

had a negative effect beyond its optimum proportion, reflecting ammonia toxicity at higher inclusion levels consistent with Section 2.3.

Response Surface Optimisation and Validation

Applying the desirability function to the fitted model identified an optimum blend of 40% FW, 25% VMW, 20% FrMW, 10% CD and 5% SHW (by VS), achieving a blended C/N ratio of 24.8:1 — within the ideal 20–30:1 range — and a predicted biogas yield of 718 mL/g VS. Independent triplicate validation trials at this composition produced 724 ± 18 mL/g VS (within 1% of the model prediction) with a methane content of 63.4%, corresponding to a specific methane yield of approximately 459 mL CH_4/g VS. Table 3 benchmarks this result against comparable co-digestion studies from the literature, showing that the five-substrate optimised blend exceeds the yields reported for simpler two- or three-substrate combinations, supporting the conclusion that multi-substrate RSM optimisation captures synergies beyond what pairwise blending alone achieves.

Table 3: Comparison of this study's optimised co-digestion yield with published co-digestion studies

Study / Substrate Combination	Blending Ratio	Yield
Elango et al. [1] – MSW + sewage	Single-factor OLR optimum	0.36 m³ biogas/ kg VS
Bouallagui et al. [5] – VMW + CD	3:1	0.41 m³ biogas/ kg VS
Singh et al. [6] – VMW + CD (Azadpur)	2:1	0.58 m³ biogas/ kg VS
Pavi et al. [7] – OFMSW + fruit/ veg. waste	1:3	396.6 mL CH_4/g VS
This study – FW + VMW + FrMW + CD + SHW	40:25:20:10:5	459 mL CH_4/g VS (724 mL biogas/g VS)

Process Stability at Design OLR and HRT

Single-factor trials on the optimised blend identified an optimum OLR of 3.0 g VS/L·day, beyond which yield declined progressively (Figure 2), and an optimum HRT of 20 days, achieving 78.4% VS removal with diminishing returns at longer retention (Figure 3). A 30-day semi-continuous trial at this operating point (Table 4) showed the FOS/TAC ratio decline from 0.28 to 0.14 and free ammonia nitrogen remain at 42–51 mg N/L, both within stable operating ranges and well below the 100–200 mg N/L inhibition threshold reported by Yeniğün and Demirel [8] — confirming that the optimised blend avoids the ammonia accumulation risk associated with SHW-rich mixtures identified during screening.

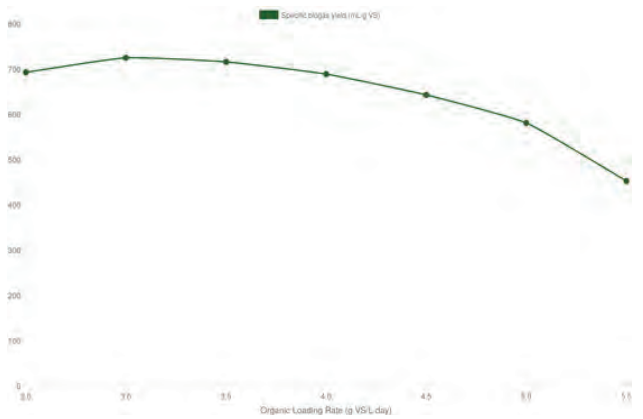


Fig. 2: Effect of organic loading rate (OLR) on specific biogas yield

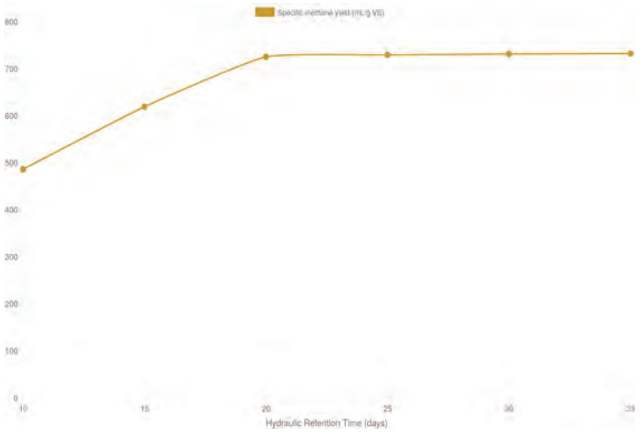


Fig. 3: Effect of hydraulic retention time (HRT) on specific methane yield

Digestate Quality

The digestate solid fraction met FCO (1985) Schedule VI biofertiliser standards for total nitrogen (4.62% dry wt.), P_2O_5 (1.68% dry wt.) and faecal coliform count, adding further value as an organic manure by-product and supporting a secondary revenue stream for the proposed plant.

Conceptual Plant Design and Techno-Economic Feasibility

Table 4 summarises the resulting plant design and techno-economic indicators for a conceptual 200 MT/day facility based on the optimised formulation. The projected internal rate of return comfortably exceeds the 10–12% cost of capital typical of Indian municipal infrastructure projects, and one-at-a-time sensitivity analysis indicated that the IRR remained above the cost-of-capital threshold across the tested range of capital cost and feedstock/tipping cost variation, indicating the project's economics are not narrowly dependent on any single favourable assumption.

Table 4: Process operating parameters and techno-economic indicators (200 MT/day plant)

Parameter	Value
Optimum OLR	3.0 g VS/L·day
Optimum HRT (VS removal)	20 days (78.4%)
Digester configuration	4 × 2,250 m³ CSTR
Gross biogas production	~27,100 m³/day
Net exportable power	~2.4 MWe
Capital cost	~Rs. 78.84 crore
Simple payback period	6.8 years
Internal rate of return (IRR)	18.4%
Discount rate used for NPV/IRR	12% over 25-year life

CONCLUSIONS

The principal contribution of this study is methodological: characterising KMC's organic waste as five distinct streams, rather than a single undifferentiated fraction, and combining a structured screening-then-optimisation experimental pipeline with RSM, exposed complementary C/N and biodegradability profiles that mono-digestion studies cannot capture. The resulting yield gain over comparable two-substrate co-digestion reported in the literature (Table 3) suggests that multi-substrate RSM optimisation captures synergies that pairwise blending misses — particularly the buffering role of cattle dung and the under-recognised yield contribution of fruit market waste. Because the operating envelope and economic outcome reported here were derived from the same waste-characterisation-to-optimisation pipeline, described in full in Section 3, the approach is directly transferable to other Indian Tier-2 cities with similarly mixed organic waste streams, offering a template for evidence-based biogas plant design in place of generic loading assumptions.

The principal limitation of this study is that its findings rest on bench-scale batch and semi-continuous trials; pilot-scale continuous validation at a scale intermediate between the bench and the proposed 200 MT/day facility remains necessary before full-scale deployment. Future work should additionally examine pre-treatment of the lignocellulosic fraction of vegetable market waste to further improve yield, agronomic field trials of the resulting digestate on local soils, and a comparative assessment of centralised versus decentralised plant configurations for KMC's ward-level waste collection network.

ACKNOWLEDGEMENTS

The authors thank the Yashwantrao Chavan School of Rural Development, Shivaji University, Kolhapur, and Kolhapur Municipal Corporation for facilitating access to waste sampling sites and laboratory facilities used in this study.

REFERENCES

1. D. Elango, M. Pulikesi, P. Baskaralingam, V. Ramamurthi, S. Sivanesan (2007), Production of biogas from municipal solid waste with domestic sewage, Journal of Hazardous Materials, Vol. 141, No. 1, pp. 301–304.
2. Wan Azlina AB Karim Ghani, Azni Idris (2009), Preliminary study on biogas production of biogas from municipal solid waste (MSW) leachate, Journal of Engineering Science and Technology, Vol. 4, No. 4, pp. 374–380.
3. Karthik Rajendran, Harshavardhan R. Kankanala, Rakel Martinsson, Mohammad J. Taherzadeh (2013), Uncertainty over techno-economic potentials of biogas from municipal solid waste (MSW), Journal of Hazardous Materials, Vol. 141, No. 1, pp. 301–304.
4. J. Mata-Alvarez, J. Dosta, M.S. Romero-Güiza, X. Fontécha, S. Peces, S. Astals (2014), A critical review on anaerobic co-digestion achievements between 2010 and 2013, Renewable and Sustainable Energy Reviews, Vol. 36, pp. 412–427.
5. H. Bouallagui, Y. Touhami, R. Ben Cheikh, M. Hamdi (2009), Bioreactor performance in anaerobic digestion

of fruit and vegetable wastes, Process Biochemistry, Vol. 40, No. 3–4, pp. 989–995.

6. A. Singh, D. Sharma, S. Kumar (2020), Co-digestion of vegetable market waste and cattle dung: A case study of Azadpur mandi, International Journal of Green Technologies, Vol. 27, No. 3, pp. 156–165.
7. Suelen Pavi, Luis Eduardo Kramer, Luciana Paulo Gomes, Luis Alcides Schiavo Miranda (2017), Biogas production from co-digestion of organic fraction of municipal solid waste and fruit and vegetable waste, Bioresource Technology, Vol. 228, pp. 361–367.
8. Özgür Yeniğün, Burak Demirel (2013), Ammonia inhibition in anaerobic digestion: A review, Process Biochemistry, Vol. 48, No. 5–6, pp. 901–911.
9. Sagarika Panigrahi, Brajesh K. Dubey (2019), A critical review on operating parameters and strategies to improve the biogas yield from anaerobic digestion of organic fraction of municipal solid waste, Renewable Energy, Vol. 143, pp. 779–797.
10. Michele Pognani, Giuliana D'Imporzano, Barbara Scaglia, Fabrizio Adani (2009), Substituting energy crops with organic fraction of municipal solid waste for biogas production at farm level, Process Biochemistry, Vol. 44, No. 8, pp. 817–821.
11. Barbara Scaglia, Roberto Confalonieri, Giuliana D'Imporzano, Fabrizio Adani (2010), Estimating biogas production of biologically treated municipal solid waste, Bioresource Technology, Vol. 101, No. 3, pp. 945–952.
12. Vanessa Abad, Romina Avila, Teresa Vicent, Xavier Font (2019), Promoting circular economy in the surroundings of an organic fraction of municipal solid waste anaerobic digestion treatment plant: Biogas production impact and economic factors, Bioresource Technology, Vol. 283, pp. 10–17.
13. Central Pollution Control Board (CPCB) (2021), Annual Report on Solid Waste Management, Ministry of Environment, Forest and Climate Change, Government of India.
14. Ministry of New and Renewable Energy (MNRE) (2022), National Bioenergy Programme – Guidelines, Government of India.
15. K. Rajendran, S. Aslanzadeh, M.J. Taherzadeh (2012), Household biogas digesters—A review, Energies, Vol. 5, No. 8, pp. 2911–2942.
16. D. Kondusamy, A.S. Kalamdhad (2014), Pre-treatment and anaerobic digestion of food waste for high rate methane production – A review, Journal of Environmental Chemical Engineering, Vol. 2, No. 3, pp. 1821–1830.
17. Y. Chen, J.J. Cheng, K.S. Creamer (2008), Inhibition of anaerobic digestion process: A review, Bioresource Technology, Vol. 99, No. 10, pp. 4044–4064.
18. I. Angelidaki, M. Alves, D. Bolzonella, L. Borzacconi, J.L. Campos, A.J. Guwy, S. Kalyuzhnyi, P. Jenicek, J.B. van Lier (2009), Defining the biomethane potential (BMP) of solid organic wastes and energy crops: a proposed protocol for batch assays, Water Science and Technology, Vol. 59, No. 5, pp. 927–934.
19. VDI 4630 (2006), Fermentation of Organic Materials – Characterisation of the Substrate, Sampling, Collection of Material Data, Fermentation Tests, Verein Deutscher Ingenieure, Germany.
20. APHA (2017), Standard Methods for the Examination of Water and Wastewater, 23rd Edition, American Public Health Association, Washington DC.
21. R.H. Myers, D.C. Montgomery, C.M. Anderson-Cook (2016), Response Surface Methodology: Process and Product Optimization Using Designed Experiments, 4th Edition, John Wiley & Sons, New York.
22. B. Deepanraj, V. Sivasubramanian, S. Jayaraj (2014), Biogas generation through anaerobic digestion process – an overview, Research Journal of Chemistry and Environment, Vol. 18, No. 5, pp. 80–94.
23. Fertiliser Control Order (FCO) (1985), Schedule VI – Specifications for Organic Fertilisers, Ministry of Agriculture and Farmers Welfare, Government of India.
24. C. Zhang, H. Su, J. Baeyens, T. Tan (2014), Reviewing the anaerobic digestion of food waste for biogas production, Renewable and Sustainable Energy Reviews, Vol. 38, pp. 383–392.
25. Ministry of Petroleum and Natural Gas (MoPNG) (2018), SATAT – Sustainable Alternative Towards Affordable Transportation Scheme Document, Government of India.
26. A. Khalid, M. Arshad, M. Anjum, T. Mahmood, L. Dawson (2011), The anaerobic digestion of solid organic waste, Waste Management, Vol. 31, No. 8, pp. 1737–1744.
27. Kolhapur Municipal Corporation (KMC) (2024), Solid Waste Management Annual Status Report, Environment Department, Kolhapur Municipal Corporation, Kolhapur.

DOI - 10.68648/ijte-q-2026-49-03-003

Design Chart Development for Sustainable Sheet Pile Retaining Structures

Ramya L
M tech Student
Dept. of Civil Engineering
JNTUH University College of Engg. Sci. and Tech.
✉ ramyalaishetty599@gmail.com

Padmavathi M
Professor
Dept. of Civil Engineering
JNTUH University College of Engg. Sci. and Tech.
✉ mpadmace@jntuh.ac.in

ABSTRACT

Practical design tools for sheet pile wall applications are now necessary due to the growing need for affordable and sustainable retaining systems. For cantilever sheet pile walls under homogenous sand, homogeneous clay, and layered soil conditions, this study creates generalized non-dimensional equations and design charts. The suggested charts make it possible to quickly estimate the maximum bending moment and embedment depth, which eliminates the need for repeated analytical computations during early design. PLAXIS 3D finite element analysis and Rankine theory analysis were used to validate the analytical predictions, which showed good agreement and confirmed the dependability of the suggested method. Additionally, embodied carbon emissions were used to evaluate the environmental performance of standard steel sheet pile walls, recycled steel sheet pile walls, and reinforced concrete retaining walls. According to the findings, recycled steel sheet pile walls provide the best structural performance at the lowest carbon emissions. The created design charts encourage sustainable alternatives for earth-retaining structures and offer engineers a straightforward and effective design tool.

KEYWORDS *: Cantilever Sheet Pile Wall, Design Charts, Buckingham π Theorem, PLAXIS 3D, Sustainability.*

INTRODUCTION

The need for safe, affordable, and environmentally friendly retaining systems for excavations, basements, tunnels, waterfront buildings, and transit projects has grown due to the quick development of urban infrastructure. Because of its great strength, ease of installation, flexibility to adapt to various soil conditions, and possibility for reuse, steel sheet pile walls are utilized extensively. Because the immersed portion of the wall creates passive resistance, cantilever sheet pile walls are especially well suited for shallow to medium excavations. The preliminary design procedure is time-consuming since conventional analytical approaches require repeated calculations for different soil qualities and groundwater conditions, even if they offer dependable solutions for calculating embedment depth and maximum bending moment. Finite element analysis demands specialist software and a substantial amount of computational work, despite providing accurate prediction of soil-structure interaction. As a result, quick preliminary design is made possible by straightforward and dependable design tools. Furthermore, the significance of sustainable keeping systems with reduced embodied carbon emissions has been brought to light by growing environmental concerns. For cantilever sheet pile walls under homogenous sand, homogeneous clay, and layered soil conditions, this study creates generalized non-dimensional equations and useful design charts. PLAXIS 3D is used to validate the suggested analytical models, and embodied carbon assessment is used to analyze the sustainability of sheet pile walls made of reinforced concrete, conventional steel, and recycled steel. For the initial design of cantilever sheet pile retaining structures, the created design charts offer an effective and environmentally friendly method.

LITERATURE REVIEW

Sheet pile retaining wall design has advanced considerably through a combination of analytical, numerical, and reliability-based investigations.

Coulomb [1] and Rankine [2] first established the classical lateral earth pressure concepts that still underlie most retaining-structure analysis today. Terzaghi [3] brought soil mechanics principles into foundation engineering practice, and this was followed by Rowe [4] and Brinch Hansen [5], who worked out analytical procedures specifically for sheet pile wall design. Broms [6] later contributed further insight into how embedded retaining systems respond to lateral loading. Much of this design knowledge has since been formalized through manuals and codes, including NAVFAC [8], the U.S. Army Corps of Engineers [9], Eurocode 7 [25], and BS 8002 [27], along with established texts by Peck et al. [7], Bowles [11], Day [18], and Das [20], which together cover earth-pressure theory, embedment-depth calculation, and structural design of retaining systems across a range of soil conditions. Growth in computational geotechnics has since made it possible to model soil–structure interaction more accurately using finite element methods: Bilgin [22] applied numerical modelling to cantilever sheet pile wall behaviour, and Rajeev et al. [26] examined their seismic response through finite element analysis. Uncertainty in soil properties and loading has also been addressed through reliability-based frameworks, notably those of Ghazavi and Eslami [21] and of Ayyub and Ebeling [24]. Zakaria and Basha [28] more recently reviewed design methods for cantilever sheet pile walls in cohesionless soils and pointed to a continuing need for simplified procedures suited to everyday engineering practice. Despite this considerable body of analytical and numerical work, comparatively little effort has been directed toward generalized non-dimensional design charts spanning homogeneous sand, homogeneous clay, and layered soil conditions together, and the embodied-carbon performance of retaining systems has likewise received limited attention. This study addresses that gap by developing generalized non-dimensional equations and practical design charts for cantilever sheet pile walls, validating them against PLAXIS 3D results, and incorporating an embodied carbon assessment to support more sustainable retaining wall design.

METHODOLOGY

Design charts for anchored sheet pile walls are widely available in standard geotechnical references, but equivalent charts for cantilever sheet pile walls are limited, so their design typically relies on iterative analytical procedures. To address this gap, three representative soil profiles were adopted to capture the range of conditions commonly encountered in practice: (a) homogeneous sand, characterized by constant unit weight and angle of internal friction; (b) homogeneous clay, characterized by constant unit weight and undrained shear strength; and (c) a non-homogeneous, layered profile with sand above the dredge line and clay below it. Each profile was analysed with and without a groundwater table to capture the effect of hydrostatic pressure on wall behaviour.

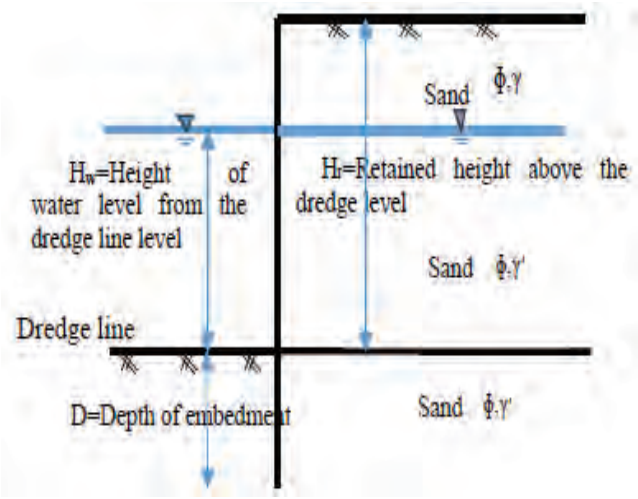


Fig. 1 Schematic cross-sections of the cantilever sheet pile wall embedded in (a) homogeneous sand

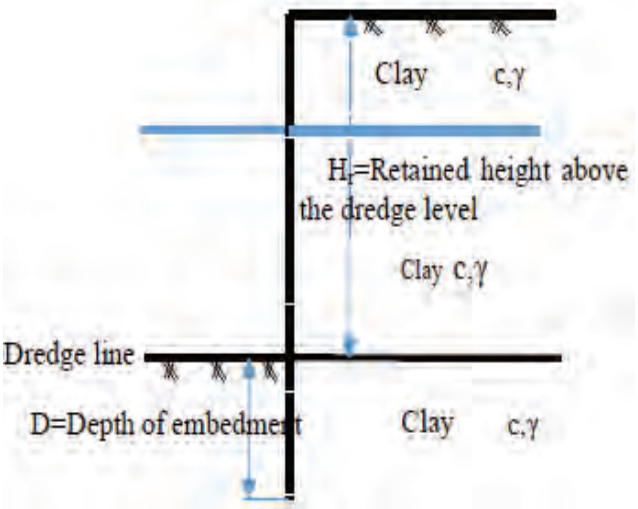


Fig.2 Schematic cross-sections of the cantilever sheet pile wall embedded in (b) homogeneous clay

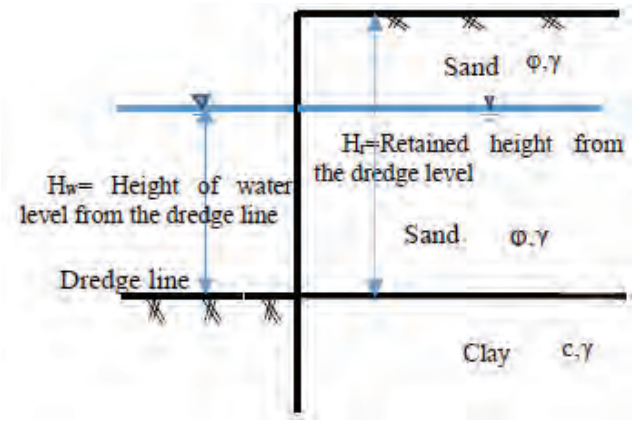


Fig.3 Schematic cross-sections of the cantilever sheet pile wall embedded in (c) layered soil

Rankine's earth-pressure theory, combined with the principle of limit equilibrium, forms the basis of the analytical formulation used here. For each soil profile, the active and passive lateral earth-pressure distributions were worked out along the wall depth and integrated to obtain the resultant forces together with their points of application.

Equilibrium of horizontal forces and of moments taken about the wall base was then applied to derive closed-form expressions for the required embedment depth (D) and the maximum bending moment (Mmax) acting on the wall. This framework provides the basis for the non-dimensionalization of the governing relationships that follows. The governing physical variables were reduced to a smaller set of independent dimensionless groups using the Buckingham π theorem, which allows the analytical results to be generalized across different soil and geometric conditions. For a phenomenon described by n governing variables involving m fundamental dimensions, it can be represented by Nπ independent dimensionless groups, Eq. (1):

$$N_{\pi} = n\text{-}m \quad (1)$$

For the homogeneous sand case, as illustrated in Fig. 1 the governing variables (D, Hr, Hw, γ, γ', φ) reduce, after non-dimensionalization, to the normalized embedment depth (D* = D/Hr), the groundwater ratio (H' = H_w/H_r), and the unit weight ratio (η = γ'/γ), as summarized

$$\pi_1 = \frac{D}{H_r} = D^*,\ \pi_2 = \frac{H_w}{H_r} = H',\ \pi_3 = \frac{\gamma'}{\gamma} = \eta$$

Several candidate functional forms linking these dimensionless groups were tested against the analytical results, and the coefficients giving the best fit were adopted, yielding the generalized design equations below. For homogeneous sand, the normalized embedment depth and normalized bending moment ($M^* = M/\gamma H r^3$) are given by Eq. (2) and (3):

$$\frac{D}{H_r} = [0.43 + 0.80(\mu^{0.20})] \times [1 + 0.32(H')(\eta^{0.20})] \quad (2)$$

$$\frac{M}{\gamma H_r^{\,3}} = \frac{0.15+0.108(\mu^{0.30})}{1+1.5H'} \quad (3)$$

Where, Hr=Retained height from the dredge level, Hw = Height of the water level from the dredge line, D/H_r = Normalized embedment depth =D^*, H_w/H_r = Groundwater level ratio =H', γ'/γ= Unit weight ratio = η, K_a / K_p = Ratio of earth coefficients = μ, $M/(\gamma H_r^3)$ = Normalized maximum bending moment =M^*

For homogeneous clay conditions shown in Fig. 2, the relevant dimensionless groups reduce to the unit weight ratio (η) and a cohesion-based stability parameter (λ = γHr/2c). The corresponding design equations for normalized embedment depth and normalized maximum bending moment are given by Eq. (4) and (5):

$$\frac{D}{H_r} = \frac{0.32\eta^{2.2}}{1+0.40\,\eta} \quad (4)$$

$$M_{max}/\gamma H_r^3 = 0.007\lambda^{2.4} / (1 + 0.75\lambda + 0.15\lambda^2) \quad (5)$$

Where, γ'/γ = Unit weight ratio=η, $(\gamma H_r)/2c$= Stability parameter for cohesive soils =λ

For the layered soil profile, as illustrated in Fig. 3 the additional cohesion parameter ($N_c = c/\gamma H_r$) and active earth-pressure coefficient (k_a) also enter the governing relationships. Eqs. (6) and (7) give the resulting normalized embedment depth and bending moment expressions for cases with a groundwater table, while Eqs. (8) and (9) give the equivalent expressions for dry conditions.

$$\frac{D}{H_r} = \frac{\left(0.74+0.25K_a^{0.5}\right)}{1+3(N_C)}\left(1+(H')(\eta)\right) \quad (6)$$

$$\frac{M}{\gamma H_r^{\,3}} = \frac{0.41\left(1+1.45K_a^{0.5}\right)}{1+4.4(N_C)} \quad (7)$$

$$\frac{D}{H_r} = \frac{1+0.65K_a{}^{0.5}}{1+3N_C} \quad (8)$$

$$\frac{M}{\gamma H_r{}^3} = \frac{0.41\left(1+1.45K_a{}^{0.5}\right)}{1+4.4(N_C)} \quad (9)$$

Corresponding equations without a groundwater table follow the same functional form with the H' term omitted, and separate coefficient sets were calibrated for each condition. The full set of non-dimensional equations for all three soil profiles, with and without groundwater, was used to generate the generalized design charts presented in the results.

The accuracy of the proposed methodology is then checked by comparing the analytical and numerical results through a percentage-error analysis. Alongside this, a sustainability assessment compares the embodied carbon emissions of reinforced concrete retaining walls, conventional steel sheet pile walls, and recycled steel sheet pile walls. Drawing on the analytical, numerical, and environmental findings together, the study closes with generalized conclusions and practical recommendations for economical, sustainable cantilever sheet pile wall design.

The total embodied carbon emissions are calculated as

$$EC = \sum_{i=1}^{n} (Q_i \times EF_i) \quad (10)$$

where

EC = Total embodied carbon emissions (kg CO_2 e)

Q_i = Quantity of the construction material

EF_i = Embodied carbon emission factor of the material (kg CO_2 e/unit)

n = Number of construction materials considered

Specifications

The soil properties used in this study were taken from standard geotechnical references and are representative of typical sand and clay deposits relevant to cantilever sheet pile wall design. The key geotechnical parameters considered are unit weight (γ), saturated unit weight (γ_{sat}), submerged unit weight (γ'), cohesion (c), angle of internal friction (ϕ), Young's modulus (E), Poisson's ratio (ν), and the active earth pressure coefficient (K_a). In addition, the retained height (H_r), groundwater level (H_w), embedment depth (D), normalized embedment depth ($D^*=D/H_r$), maximum bending moment (M), normalized bending moment ($M^* = M/\gamma H_r^3$), active earth pressure coefficient (k_a), normalized cohesion parameter ($C/\gamma H_r$), and non-dimensional cohesion parameter ($N_c=\gamma H_r/2C$)were all fed into the analytical formulation and the Buckingham π theorem to derive the generalized non-dimensional equations and design charts. Together, these parameters control the active and passive earth pressures, embedment depth, bending moment, and overall stability of cantilever sheet pile walls. The same soil properties and parameters were used in both the analytical calculations and the PLAXIS 3D numerical models, which adopted the Mohr–Coulomb constitutive model to give a realistic representation of soil behaviour and a dependable basis for the generalized design equations and charts that follow.

PROPOSED FRAMEWORK

Figs. 1–3 illustrate the overall framework followed in this study. Work begins with a literature review used to pin down the research gap and objectives, followed by analytical investigations of homogeneous sand, homogeneous clay, and layered soil under both dry and groundwater conditions, applying Rankine and Coulomb earth-pressure theories. A parametric study then examines how the governing geotechnical parameters affect embedment depth, sheet pile length, and maximum bending moment for cantilever sheet pile walls.

The Buckingham π theorem is then used to derive generalized non-dimensional equations and matching design charts for the normalized embedment depth and normalized maximum bending moment. This methodology is checked against PLAXIS 3D finite element results, with the analytical and numerical outputs compared through a percentage-error analysis. The study concludes with a sustainability assessment comparing reinforced concrete, conventional steel, and recycled steel sheet pile walls, from which practical recommendations for sustainable cantilever sheet pile wall design are drawn.

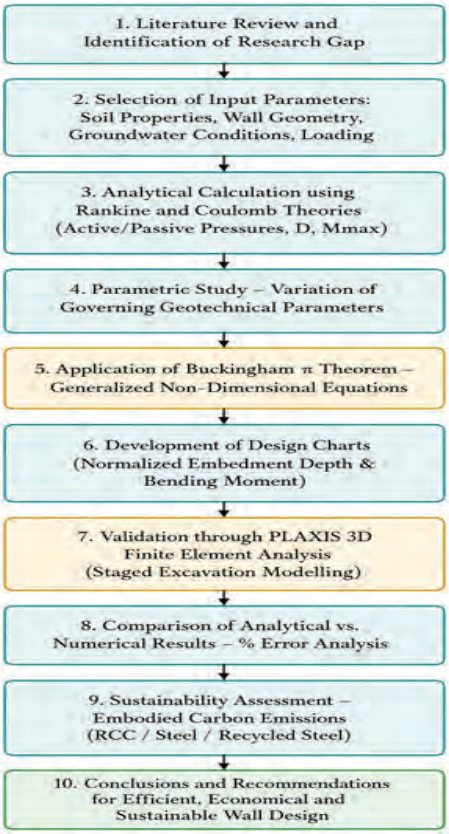


Fig. 4 Proposed framework flow chart

RESULTS AND DISCUSSION

Validation of Analytical Predictions using PLAXIS 3D

Validation of the proposed non-dimensional equations and design charts was carried out using three-dimensional finite element analysis in PLAXIS 3D, with numerical simulations run for selected cases of homogeneous sand, homogeneous clay, and layered soil, using the same soil parameters, sheet pile properties, and groundwater conditions as the analytical study. Soil behaviour was represented using the Mohr–Coulomb model, and the sheet pile wall itself was modelled with plate elements assigned appropriate structural properties.

Table 1 summarizes the comparison between the numerical and analytical values of embedment depth and maximum bending moment, along with the percentage error used to gauge the accuracy of the developed relationships. Analytical and numerical solutions agreed well overall, with errors averaging roughly 5–14% across the cases examined, which supports the reliability of the proposed design charts for cantilever sheet pile walls across the soil and water-table conditions considered. On this basis, the generalized design charts presented in the following subsections can be applied with reasonable confidence.

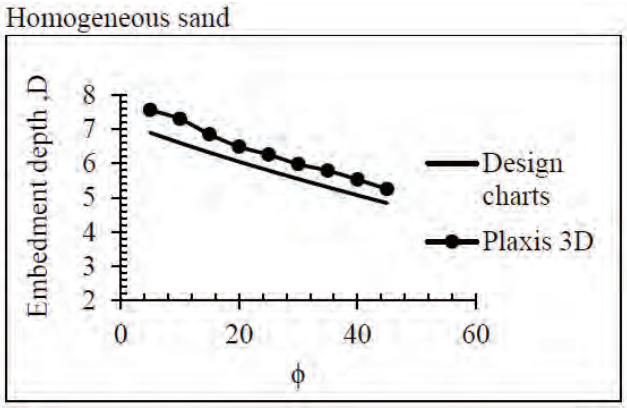


Fig. 5 Comparison of Design Charts and PLAXIS 3D Results for D vs ϕ

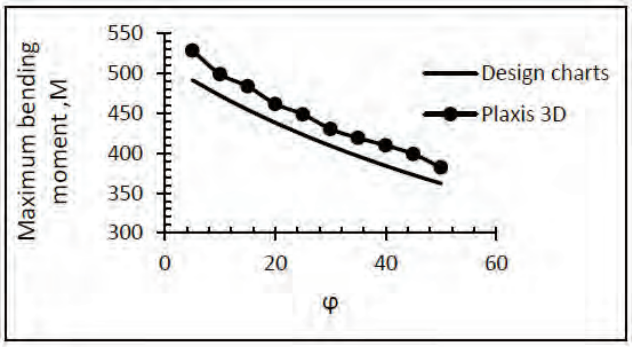


Fig. 6 Comparison of Design Charts and PLAXIS 3D Results for M vs ϕ

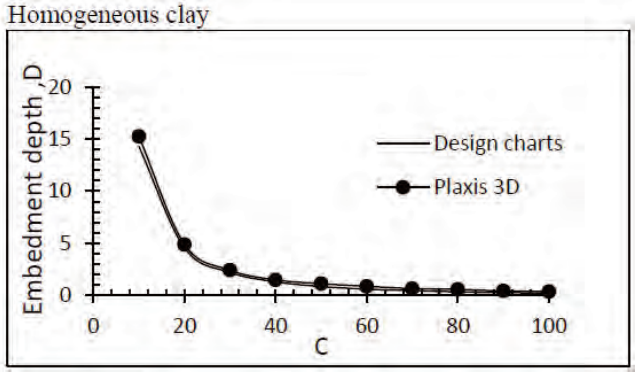


Fig. 7 Comparison of Design Charts and PLAXIS 3D Results for D vs Cohesion

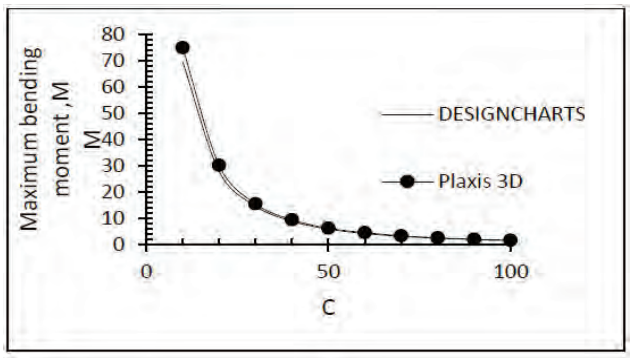


Fig. 8 Comparison of Design Charts and PLAXIS 3D Results for M vs Cohesion

Layered soil

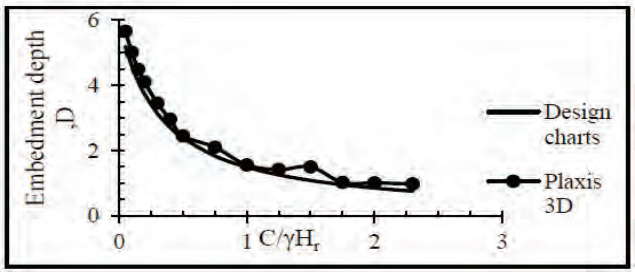


Fig. 9 Comparison of Design Chart and PLAXIS 3D Results for c/(γH$_r$)vs. D under Groundwater Conditions

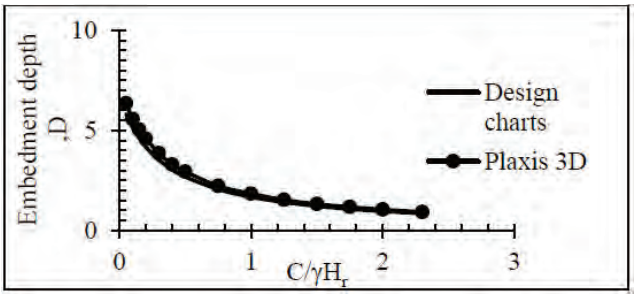


Fig. 10 Comparison of Design Chart and PLAXIS 3D Results for c/(γH$_r$)vs. D*under Dry Conditions

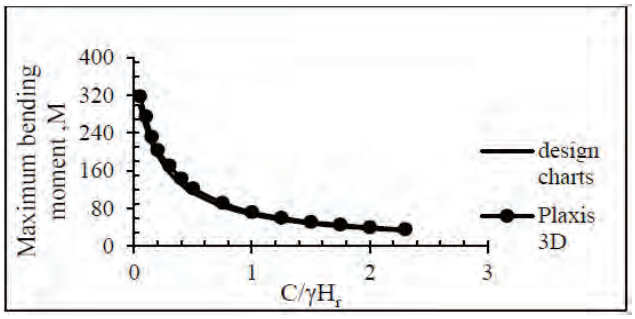


Fig. 11 Comparison of Design Charts and PLAXIS 3D Results for c/(γH$_r$) vs M under Groundwater Conditions

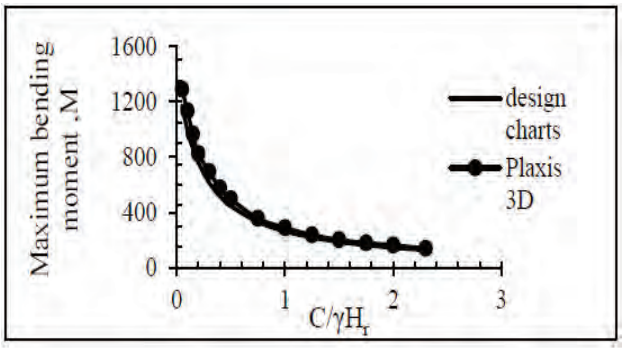


Fig. 12 Comparison of Design Charts and PLAXIS 3D Results for c/(γH$_r$) vs M under Groundwater Conditions

Table 1. Comparison of design-chart predictions with PLAXIS 3D simulations

Case	Comparison	Average Error (%)	Error Range (%)
Homogeneous sand	Embedment depth (D) vs φ	8.68	7.25 – 10.59
Homogeneous sand	Bending moment (M) vs φ	6.11	5.12 – 7.52
Homogeneous clay	Bending moment (M) vs cohesion	7.22	5.86 – 9.11
Homogeneous clay	Embedment depth (D) vs cohesion	9.34	6.30 – 13.04
Layered soil (with water table)	D vs c/γHr	10.40	2.51 – 27.62
Layered soil (without water table)	D vs c/γHr	9.04	8.06 – 9.98
Layered soil (with water table)	M vs c/γHr	8.47	5.80 – 10.18
Layered soil (without water table)	M vs c/γHr	9.91	7.13 – 13.95

The proposed non-dimensional design charts are validated in Figures 5–12, which set the analytical predictions against PLAXIS 3D results for three representative cases: homogeneous sand (Figs. 5–6), homogeneous clay (Figs. 7–8), and layered soil, sand over clay, examined both with (Figs. 9–10) and without (Figs. 11–12) a groundwater table. Table 1 shows the average prediction error falling between 6.11%

and 10.40%, with most cases below 10%, indicating close agreement between the analytical and PLAXIS 3D solutions. This confirms that the proposed non-dimensional equations and design charts can reliably estimate embedment depth and maximum bending moment for cantilever sheet pile walls across the soil and groundwater conditions studied.

Design Charts for Cantilever Sheet Pile Walls

Using the Buckingham π theorem, generalized non-dimensional equations were derived for cantilever sheet pile walls in homogeneous sand, converting the analytical results into dimensionless relationships for the normalized embedment depth (D/H) and normalized maximum bending moment (Mmax/γHr

These non-dimensional equations were then used to build practical design charts for homogeneous sand, allowing the normalized embedment depth and normalized maximum bending moment to be estimated quickly without repeating the analytical calculations each time.

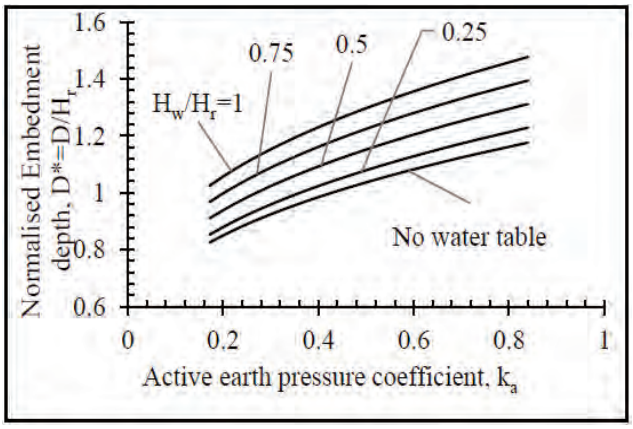


Fig. 13 Variation of normalized embedment depth, D* with active earth pressure coefficient, k_a

Figure 13 presents the design chart relating normalized embedment depth (D/H_r) to the active earth pressure coefficient (K_a) under varying groundwater conditions, based on the proposed relationship given in Eq. (2). As K_a increases from 0.2 to 0.8, D/H_r rises by roughly 50%, and a higher groundwater level pushes the required embedment depth up further, by around 20–25%. This chart therefore serves as a quick design tool for cantilever sheet pile walls across different groundwater conditions.

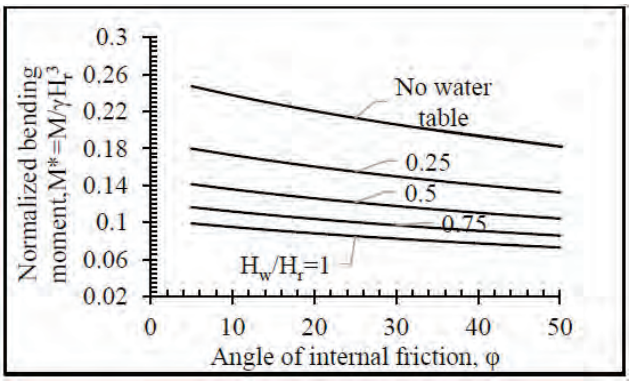


Fig. 14 Variation of normalized bending moment, M^* with angle of internal friction angle, φ

Figure 14 presents the design chart relating the normalized bending moment (M^*) to the angle of internal friction (ϕ), developed using the non-dimensional relationship given in Eq. (3). As ϕ increases from 5° to 50°, the normalized bending moment decreases by approximately 60–70%, while increasing the groundwater ratio (H_w/H_r)from 0 to 1.0 further reduces M^* by around 55–65%. This makes the chart a fast and dependable means of estimating the normalized bending moment of cantilever sheet pile walls under varying groundwater conditions.

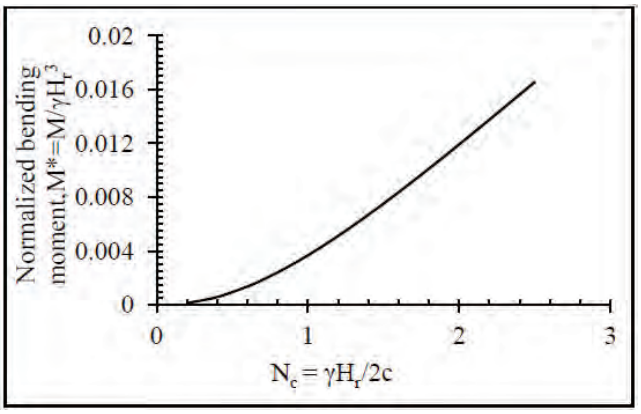


Fig. 15 Variation of the normalized bending moment, M^* with the soil strength parameter, $N_c = (\gamma H_r)/2c$

Figure 15 presents the variation of normalized bending moment ($M^* = M/\gamma H_r^3$) with the non-dimensional cohesion parameter ($N_c = \gamma H_r/2C$) for homogeneous clay under dry conditions, based on the proposed relationship given in Eq. (4). The chart shows that M^* increases with increasing N_c, indicating reduced cohesion influence and higher bending demand. For

example, M^* increases from approximately 0.008 at N_c = 1.5 to 0.016 at N_c = 2.5. The developed design chart provides a simple approach for estimating bending moments in cantilever sheet pile walls.

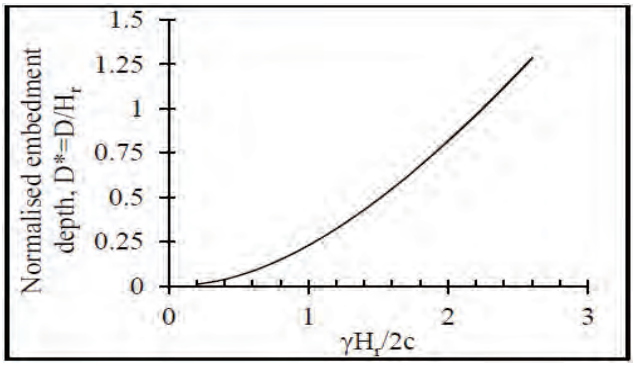


Fig. 16 Variation of the normalized depth ratio D* with the normalized soil parameter (γH_r)/2c

Figure 16 presents the variation of normalized embedment depth ($D^* = D/H_r$)with the non-dimensional cohesion parameter ($N_c = \gamma H_r/2C$)for homogeneous clay under dry conditions, based on Eq. (5). The results show that D^*increases with N_c, rising from approximately 0.50 at N_c = 1.5 to 1.30 at N_c =2.6, reflecting the greater embedment needed as the influence of cohesion diminishes. The proposed chart offers a dependable way of estimating embedment depth for cantilever sheet pile walls.

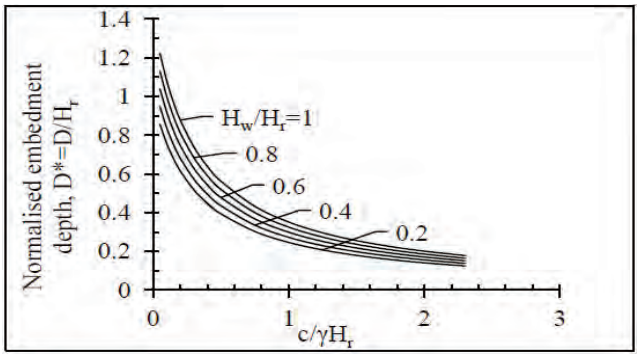


Fig. 17 Variation of the normalized cohesion parameter with the normalized depth ratio D* in the absence of water table

Figure 17 presents the variation of normalized embedment depth ($D^* = D/H_r$) with the normalized cohesion parameter ($C/\gamma H_r$) for non-homogeneous soil under different groundwater levels, based on the relationship given in Eq. (6). The results show that D^* decreases with increasing cohesion, whereas higher groundwater levels increase the embedment requirement. At $C/\gamma H_r$ = 1.0, D^* increases from 0.22 to 0.35 as H_w/H_{rrises} from 0.2 to 1.0. This design chart offers an efficient means of estimating embedment depth for cantilever sheet pile walls under varying groundwater conditions.

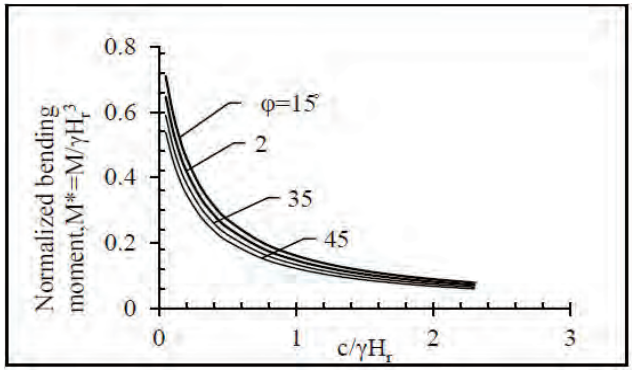


Fig. 18 Variation of the normalized cohesion parameter c/(γH_r) with the non-dimensional moment parameter, M* in the absence of water table

Figure18 presents the variation of normalized bending moment ($M^* = M/\gamma H_r^3$)with the normalized cohesion parameter($C/\gamma H_r$) for non-homogeneous soil at different friction angles without groundwater, following the relationship in Eq. (7). The results indicate that M^* decreases with increasing cohesion and friction angle due to enhanced soil strength and reduced lateral earth pressure. At $C/\gamma H_r$ = 1.0, M^* decreases from 0.17 to 0.11 as ϕ rises from 15° to 45°. The chart therefore gives a dependable tool for estimating bending moments in cantilever sheet pile walls under dry conditions.

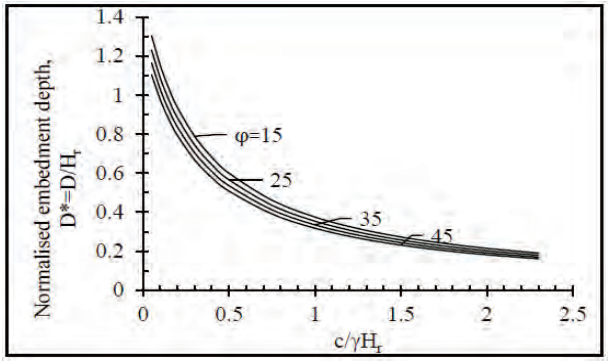


Fig. 19 Variation of the normalized cohesion parameter c/(γH_r)with the normalized depth ratio D* in the presence of water table

Figure 19 presents the variation of normalized embedment depth ($D^* = D/H_r$)with the normalized cohesion parameter ($C/\gamma H_r$) for non-homogeneous soil at different friction angles with groundwater present, following the relationship in Eq. (8). The results show that D^* decreases with increasing cohesion and friction angle due to enhanced soil resistance. At $C/\gamma H_r = 1.0$, D^* reduces from 0.35 to 0.29 as ϕ increases from 15° to 45°. The proposed chart provides a reliable approach for estimating embedment depth of cantilever sheet pile walls.

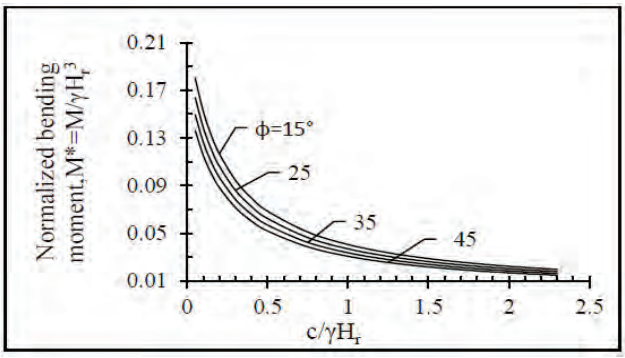


Fig. 20 Variation of the normalized cohesion parameter c/(γH_r) with the non-dimensional moment parameter M* in the presence of water table

Figure 20 presents the variation of normalized bending moment ($M^* = M/\gamma H_r^3$)with the normalized cohesion parameter ($C/\gamma H_r$) for non-homogeneous soil at different friction angles with groundwater present, following the relationship in Eq. (9). The results show that M^* decreases with increasing cohesion and friction angle due to reduced lateral earth pressure. At $C/\gamma H_r = 1.0$, M^* reduces from 0.038 to 0.029 as ϕincreases from 15° to 45°. The proposed chart provides a reliable approach for estimating bending moments of cantilever sheet pile walls.

Figure 21 sets out the CO_2 emissions of the three retaining wall systems for a 1 m wall length. The RCC retaining wall generates the most emissions (3765 kg CO_2-e), the conventional steel sheet pile wall comes next (2736 kg CO_2-e), and the recycled steel sheet pile wall produces the least (864 kg CO_2-e). Switching to recycled steel therefore cuts CO_2 emissions by around 68% relative to conventional steel and by 77% relative to the RCC wall, underlining its stronger environmental performance.

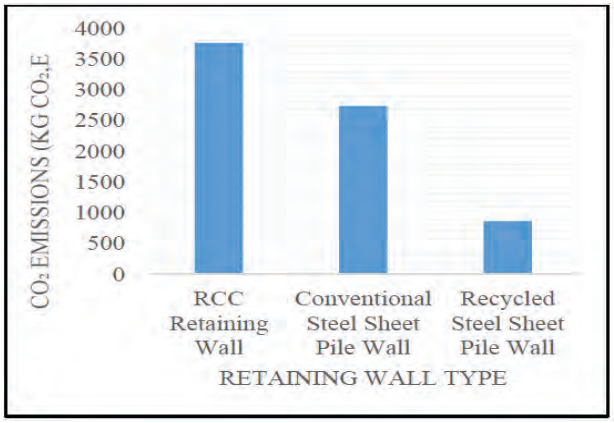


Fig. 21 Comparison of CO_2 Emissions of Different Retaining Wall Systems

CONCLUSIONS

The design charts for homogeneous sand showed that the normalized embedment depth increased from 0.82 to 1.48 with an increase in K_a from 0.2 to 0.8, while the normalized bending moment fell by 25–30% as the friction angle increased. Both parameters rose under groundwater conditions because of hydrostatic pressure effects. For homogeneous clay, the parameter $N_c=\gamma H_r/2c$ significantly influenced wall behaviour, with an increase in N_c rising from 0.25 to 2.5 pushed the normalized embedment depth up from 0.02 to 1.28 and the bending moment from 0.0002 to 0.016. The layered soil design charts successfully captured the effects of soil stratification and groundwater conditions. Increasing $c/\gamma H_r$ brought down both the normalized embedment depth and bending moment, while the presence of groundwater increased the wall's response. PLAXIS 3D simulations validated the proposed analytical equations, with errors ranging from 7.1% to 14.0% and averaging around 10%, supporting the reliability of the methodology developed. The sustainability assessment further showed that recycled steel sheet piles cut embodied carbon emissions by 68% relative to conventional steel sheet piles and by 77% relative to RCC retaining walls, pointing to their potential as a more sustainable retaining option.

ACKNOWLEDGEMENT

The authors gratefully thank Dr. M. Padmavathi, Professor, Department of Civil Engineering, University College of Engineering Science & Technology

Hyderabad (UCESTH), JNTUH, for her guidance, encouragement, and constructive input throughout this work. The authors also thank UCESTH, JNTUH, for the academic facilities and support extended during the course of this study.

REFERENCES

1. C. A. Coulomb, "Essay on the Application of the Rules of Maxima and Minima to Some Problems of Statics Related to Architecture," Memoirs of the Royal Academy of Sciences, Paris, France, 1776.
2. W. J. M. Rankine, "On the Stability of Loose Earth," Philosophical Transactions of the Royal Society of London, vol. 147, pp. 9–27, 1857.
3. K. Terzaghi, Theoretical Soil Mechanics. New York, NY, USA: John Wiley & Sons, 1943.
4. P. W. Rowe, "Anchored Sheet-Pile Walls," Proceedings of the Institution of Civil Engineers, vol. 1, no. 1, pp. 27–70, 1952.
5. J. Brinch Hansen, "Earth Pressure Calculation," The Danish Geotechnical Institute Bulletin, Copenhagen, Denmark, 1953.
6. B. B. Broms, "Lateral Resistance of Piles in Cohesive Soils," Journal of the Soil Mechanics and Foundations Division, ASCE, vol. 90, no. SM2, pp. 27–63, 1964.
7. R. B. Peck, W. E. Hanson, and T. H. Thornburn, Foundation Engineering, 2nd ed. New York, NY, USA: John Wiley & Sons, 1974.
8. U.S. Department of the Navy, Foundations and Earth Structures, NAVFAC Design Manual DM 7.02. Alexandria, VA, USA: Naval Facilities Engineering Command, 1986.
9. U.S. Army Corps of Engineers, Engineering Manual EM 1110-2-2504: Design of Sheet Pile Walls. Washington, DC, USA: Department of the Army, 1994.
10. U.S. Army Corps of Engineers, Engineering and Design—Design of Sheet Pile Walls, EM 1110-2-2504, Washington, DC, USA: Department of the Army, Jul. 1994.
11. J. E. Bowles, Foundation Analysis and Design, 5th ed. New York, NY, USA: McGraw-Hill, 1996.
12. Naval Facilities Engineering Command (NAVFAC), Foundations and Earth Structures, Design Manual DM-7.02, Alexandria, VA, USA: NAVFAC, 1986.
13. Bureau of Indian Standards, IS 14458 (Part 1): 1998—Guidelines for Retaining Wall for Hill Areas: Selection of Type of Wall. New Delhi, India: BIS, 1998.
14. Bureau of Indian Standards, IS 14458 (Part 2): 1997—Guidelines for Retaining Wall for Hill Areas: Design of Retaining/Breast Walls. New Delhi, India: BIS, 1997.
15. Bureau of Indian Standards, IS 14458 (Part 3): 1998—Guidelines for Retaining Wall for Hill Areas: Construction of Dry Stone Walls. New Delhi, India: BIS, 1998.
16. Bureau of Indian Standards, IS 14458 (Part 4): 1998—Guidelines for Retaining Wall for Hill Areas: Construction of Retaining and Breast Walls. New Delhi, India: BIS, 1998
17. W. Powrie, Soil Mechanics: Concepts and Applications. London, U.K.: E & FN Spon, 1996.
18. R. W. Day, Geotechnical and Foundation Engineering: Design and Construction. New York, NY, USA: McGraw-Hill, 1999.
19. R. W. Day, Foundation Engineering Handbook: Design and Construction with the 2001 International Building Code. New York, NY, USA: McGraw-Hill, 2001.
20. B. M. Das, Principles of Foundation Engineering, 6th ed. Belmont, CA, USA: Thomson Brooks/Cole, 2007.
21. M. Ghazavi and A. Eslami, "Reliability-Based Design of Cantilever Sheet Pile Walls in Sandy Soils," International Journal of Civil Engineering, vol. 6, no. 4, pp. 257–268, 2008.
22. Ö. Bilgin, "Numerical Studies of Cantilever Sheet Pile Walls," Computers and Geotechnics, vol. 37, no. 3, pp. 399–407, 2010.
23. D. P. Coduto, M. R. Yeung, and W. A. Kitch, Geotechnical Engineering: Principles and Practices, 2nd ed. Upper Saddle River, NJ, USA: Pearson, 2011.
24. B. M. Ayyub and R. M. Ebeling, Reliability-Based Design of Sheet Pile Structures. Reston, VA, USA: ASCE Press, 2011.
25. British Standards Institution, BS EN 1997-1:2004+A1:2013 Eurocode 7: Geotechnical Design—Part 1: General Rules. London, U.K.: BSI, 2013
26. P. Rajeev, Q. Bui, and N. Sivakugan, "Seismic Behaviour of Sheet Pile Retaining Walls Using Finite Element Analysis," Soil Dynamics and Earthquake Engineering, vol. 75, pp. 115–126, 2015.
27. British Standards Institution, BS 8002:2015—Code of Practice for Earth Retaining Structures. London, U.K.: BSI, 2015.
28. M. Zakaria and M. Basha, "A Review on Cantilever Sheet Pile Wall Design in Cohesionless Soils," Journal of Geotechnical and Geoenvironmental Engineering, vol. 150, no. 2, 2024.

DOI - 10.68648/ijte-q-2026-49-03-004

Feasibility of Collaborative Robot-Assisted Wall Laying Using Interlocking Compressed Earth Blocks

Vishal Kumar Chaudhary
Ph.D. Scholar
Department of Civil and Environmental Engineering
IIT Patna
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Khagaria
✉ vishal_2522ce09@iitp.ac.in

Hemant Kumar Prajapati
Senior Structural Engineer
Jacobs
✉ hkumar.hk5@gmail.com

ABSTRACT

Masonry laying remains one of the least mechanised construction trades, while the interlocking compressed earth block (ICEB) remains one of the least mechanised low-carbon walling materials at the placement stage. The paper argues the two problems solve each other: the properties masonry robotics has had to engineer at cost (mortarless assembly, self-aligning interlock geometry, dimensional uniformity and low unit mass) already exist in the ICEB, a unit codified by Indian Standards. A framework for a collaborative laying cell is proposed, comprising a block feed, a force-limited arm, two-camera perception and a wall-plan sequencer, with the mason retaining corners, openings, reinforcement and grouting. Four feasibility results follow from codal values and published measurements: a lead-in clearance of 6.9 mm absorbs the codal block tolerance (±2 mm) and demonstrated cell accuracy (2.90 mm), while course accumulation requires per-batch calibration of the effective block dimension; code-derived unit masses of 1.2–9.9 kg sit within 12–16 kg collaborative-arm payloads; and the laying rate reaches 70–130 blocks per hour with relocations amortised. Deployment is constrained by block supply and by the infill fraction of the wall, not by arm speed. The framework is analytical and awaits physical validation, for which success criteria are stated.

***KEYWORDS** : Interlocking compressed earth block; Construction robotics; Dry-stack masonry; Sustainable construction.*

NOTATIONS

e_k : cell placement error (mm)

t_b : codal block tolerance (mm)

c_{li} : interlock lead-in clearance (mm)

Δ_N : end-of-course deviation over N blocks (mm)

t_{sys}, t_{ran} : systematic and random tolerance components (mm)

m : unit mass (kg)

ρ_d : dry density (kg/m³)

V : block volume (m³)

T_c : cycle time (s)

t_p, t_g, t_m, t_v : pick, gate, move and place components (s)

Q : throughput (blocks/h)

r_e : arm reach (m)

d : wall stand-off (m)

L_{st} : usable frontage per station (m).

INTRODUCTION

Construction productivity has grown by about one percent annually over two decades, against 2.8 percent for the world economy and 3.6 percent for manufacturing [1]. Masonry wall laying remains among the least mechanised site trades even as the industry, worth close to USD 10 trillion per annum, faces shortages of skilled workers [2], and its materials are carbon-intensive: reinforced cement concrete demands

1.69 times more energy than a compressed stabilised earth block system and emits 1.64 times more CO_2-equivalent than an interlocking-block system [3].

Interlocking compressed earth blocks (ICEBs) are an established low-cost, low-carbon walling technology for social housing, though build-quality variability and skill dependence limit wider use [4]. In India IS 1725:2023 fixes classes, sizes and tolerances for the stabilized soil block [5], and IS 17165:2020 gives manufacturing guidelines recognising manual, semi-automatic and fully automatic block-making machines [6]. Production of the unit is thus on an automation trajectory; placement is not.

Interlocking units with robotic laying are not themselves unexplored: research systems have adopted purpose-designed interlocking blocks, and simulation found modular construction with interlocking blocks and robotic handling gave the greatest productivity improvement of all scenarios tested [7]. The earth-block case is unexplored. A Scopus search (August 2026) returned 131 records for masonry robotics and 51 for interlocking compressed earth blocks, but only two at their intersection, neither addressing robotic placement of ICEBs, one being an agenda-setting paper that does not consider robotic unit assembly [8]. Since Scopus indexes neither patents nor industry literature, this evidences a gap in the research record rather than proving none exists. Robot-friendly interlocking units have so far been purpose-fabricated, whereas the ICEB is codified, soil-pressed and low-carbon.

The gap conceals a match: mortarless assembly, self-aligning interlock geometry, dimensional uniformity and robot-compatible mass are properties masonry robotics has repeatedly engineered at cost, and the ICEB possesses all four because its geometry was conceived so unskilled builders could lay accurate mortarless walls. The contributions are a matching of automation barriers to a codified earth-block unit, a framework for a collaborative laying cell, four feasibility results, and demonstrator requirements.

LITERATURE REVIEW

Robotic Masonry with Conventional Units

On-site bricklaying robotics dates to the mobile robot of Pritschow et al., which automated thin-bed mortar application, handled blocks up to 50 kg, and needed a device measuring block lengths to 1 mm accuracy solely to compensate unit tolerances [9]. Yu et al. verified a laying robot analytically, naming the inequality of construction materials as a principal obstacle [10], and Dakhli and Lafhaj designed one around the thin-joint method, which demands blocks rectified to 1 mm, with units weighing 17 kg [11]. Cooperative multi-robot systems have built scaffold-free masonry arches [12]. The SAM100 lays 2,000–3,000 bricks per day but still requires a human alongside to smooth the mortar before further bricks are placed, while the truck-mounted Hadrian X lays 1,000 bricks per hour fully automatically from a 3D CAD design, mortar or adhesive being delivered under pressure to its laying head [2]. An AR-guided relocatable cobot recently laid two 3 m walls of 138 wooden blocks each across 7 and 9 relocations, holding global deviations of up to 2.90 ± 2.05 mm at roughly 105 blocks per hour with speed capped at 120 mm/s [13]. The recurring cost drivers are mortar handling, unit tolerance, and unit mass driving machine scale.

Interlocking and Dry-Stack Robotic Assembly

Research has responded by redesigning the unit. Zhang et al. developed engineered interlocking blocks needing translation only, reduced surface contact giving better error tolerance while mortarless construction permits disassembly [14]. Loing et al. found topologically interlocking assemblies sensitive to production and assembly error, concluded tolerance accommodation is crucial, and proposed guiding protrusions aligned with the assembly direction [1]. Robotic dry-stacking of natural stones needs stability-based sequence planning, millimetric displacements being especially detrimental [15]. Simulation confirms the direction, the modular intervention alone cutting lead time by 46 percent [7]. The field converges on dry-stacked interlocking units but implements them as bespoke blocks.

Automation of Earthen Construction

Earthen construction is automated along wet and monolithic paths: extrusion 3D printing of earth is feasible once rheology is adjusted [16], cob printing has required bespoke extruder hardware [17], and robotic rammed earth transfers compaction to the robot [18], a route found environmentally superior to manual work

at 28 percent lower global warming potential [19]. An agenda-setting review, noting that about 30 percent of the world population lives or works in earth buildings, identifies digital production as the pathway yet does not consider robotic unit assembly [8]. The block-unit pathway, with processing completed upstream in the press and the robot only placing, remains unexamined.

Interlocking Compressed Earth Blocks

The ICEB knowledge base is mature. Sturm et al. characterised dry-stack ICEBs at a mean unit compressive strength of 2.34 MPa, modified triplet tests giving friction coefficient 0.73 and a non-zero initial shear strength of 0.035 MPa attributed to the interlock, the indentations breaking rather than slipping [20]. ICEB walls with grouted, reinforced cells reach ultimate lateral resistances of about 48–50 kN in a solid 1.8 m square specimen, falling to 13–15 kN when the wall width is halved, a capacity deriving from the reinforcement rather than the dry-stacked units [21]. The Float RAM press yields about 30 blocks per hour with three workers at unit masses of 6.4–6.8 kg [22], ultrasonic pulse velocity predicts properties of Indian hollow interlocking blocks to fair accuracy [23], and geometry-corrected strength protocols are established [24]. IS 1725:2023 fixes unit tolerance at ±2 mm, defines strength classes of 3.5–10.0 N/mm², requires dry density of at least 1750 kg/m³, and lists five modular sizes from 190 × 90 × 40 to 240 × 240 × 90 mm [5].

Human–Robot Collaboration and Perception

Safety analysis of construction robots distinguishes physical risks, attentional costs and psychological impacts [25]. Marker-less deep-network pose estimation requiring only an on-site camera system is demonstrated practice [26]. Acceptance research envisions robotic assistants performing the physically demanding repetitive work while construction personnel focus on the cognitive tasks needed to plan and supervise, and notes that job-security concerns may be alleviated where robotics is perceived as upgrading manual work to more value-added work [27].

METHODOLOGY

The study follows the analytical feasibility-verification approach established for masonry robotics [10], [12]: a framework is synthesised and checked by closed-form analyses whose every input is a codal value or published measurement, with no laboratory or field data generated. Inputs come from IS 1725:2023 [5], IS 17165:2020 [6], the relocatable cobot demonstration [13], and published ICEB production rates, masses and mechanics [20], [22], [23], at worst-case values. The application is bounded to single-storey load-bearing dwelling walls of one leaf, 140–240 mm thick, in modular interlocking units to IS 1725.

(a) Tolerance budget. A block is seated correctly if the placement error of the cell and the dimensional variation of the unit are absorbed by the interlock lead-in clearance:

$$e_k + 2t_b \le C_{li} \qquad (1)$$

where e_k is the cell-level placement error, t_b the codal block tolerance and C_{li} the total lead-in clearance of the key–recess pair.

Dry-stack courses have no mortar joint in which to absorb drift, so unit tolerances accumulate along the course in a manner that depends on their origin. Blocks pressed in one mould share its dimension, so mould oversize or wear is a systematic bias accumulating linearly, whereas ejection, moisture and handling variation is random and accumulates as a square root. For N blocks laid to a datum the end-of-course deviation is

$$\Delta_N = N \cdot t_{sys} + t_{ran}\sqrt{N} \qquad (2)$$

where t_{sys} and t_{ran} are the systematic and random components of the codal envelope. The systematic term dominates and, unlike the random term, is measurable and correctable.

(b) Payload. Unit mass is computed from codal geometry and density,

$$m = \rho_d \cdot V \qquad (3)$$

and compared with collaborative-arm payload net of gripper mass.

(c) Cycle time and productivity. The per-block cycle is the sum of pick, grasp-and-gate, move, and place-and-verify components, giving hourly throughput:

$$T_c = t_p + t_g + t_m + t_v \qquad (4)$$

$$Q = 3600 / T_c \qquad (5)$$

(d) Workspace. The usable wall frontage per station of an arm of reach r_e standing off the wall by d is

$$L_{st} = 2\sqrt{(r_e^2 - d^2)} \quad (6)$$

The laying rate is reported both steady-state and effective with relocations amortised. Economic viability is discussed qualitatively rather than modelled, since cell prices and wage rates lack the traceability of the codal inputs. The claim defended is feasibility and unit suitability with demonstrator requirements, not measured performance, and is most sensitive to the systematic tolerance component governing Eq. (2).

PROPOSED FRAMEWORK

The Matching Argument

Table 1 aligns each documented barrier of masonry automation with the ICEB property that removes it. The correspondence is not coincidental: the ICEB was designed to make dry-stack walling foolproof for unskilled builders, and the same properties make it tractable for robots. The remedy proposed after tolerance failure, guiding protrusions aligned with the assembly direction [1], describes the ICEB key that already exists; and the interlock is not cosmetic, triplet tests showing the keys break rather than slip [20].

Table 1 Masonry-automation barriers and the ICEB properties that remove them

Documented barrier	ICEB property	Sources
Mortar dosing, spreading and cure control	Dry-stack assembly; no mortar in the process	[9], [7]
Placement precision borne by machine alone	Self-aligning key–recess lead-in; translation-only seating	[1], [14]
Unit dimensional variability (1 mm rectification demanded)	Press-moulded units; codal tolerance ±2 mm	[9], [11], [5]
Unit mass drives machine scale (17–50 kg in cited systems)	1.2–9.9 kg codal units within cobot payload	[9], [11], [22], [5]
In-place process burden (extrusion rheology; ramming energy)	All material processing upstream in the press	[16], [18], [6]
Wet-process cure-time coupling	Course rise limited only by robot cycle	[7], [16]
End-of-life demolition waste	Dry joints permit robotic disassembly and reuse	[14], [8]

Fig. 1 defines the geometry the analysis operates on: the key–recess pair, the chamfered lead-in funnelling an approaching block into position, and the clearance absorbing cell placement error and unit dimensional variation.

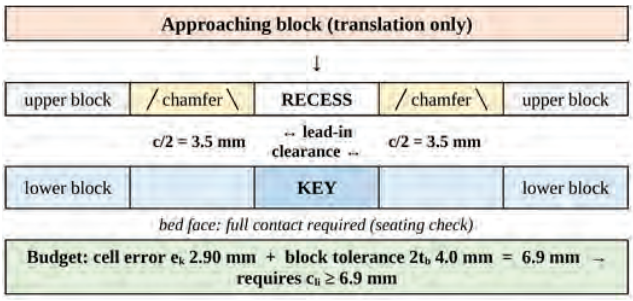


Fig. 1 Dimensioned section through a key–recess pair, showing the chamfered lead-in and the clearance budget of Eq. (1). Not to scale

Cell Architecture

Fig. 2 shows the cell. Blocks enter under three conditions rather than as free variables: produced from soil controlled to the 12–15 percent clay content IS 17165 identifies as optimum [6], since mix control upstream of the press determines how often the ±2 mm tolerance is met; cured to the requisite wet strength [6]; and laid in a stable moisture condition, since IS 1725 recognises linear expansion on saturation [5] and a mortarless wall has no joint to accommodate such movement.

Cured stock is presented at a fixed pick pose on a feed table. A collaborative arm of the 10–16 kg payload class, on a wheeled base relocated station-by-station by the mason [13], carries a force-limited parallel gripper whose wide flat jaws grasp only the plane faces, never the keys, respecting the green-strength handling

provision of IS 17165 [6]; the chipping tolerated by IS 1725 clause 5.2 is an acceptance concession for delivery handling [5], not a design allowance.

Perception is split between two viewpoints, since one fixed camera cannot serve both a close pick pose and millimetre-scale course geometry across a relocating cell. A wrist camera performs pose estimation at the pick point [26] with a defect gate applying IS 1725 clause 5.2 [5], optionally supplemented by ultrasonic pulse velocity [23]; a mast camera, re-registered at each station, verifies course level, plumb and bond and supplies the course-end position used to correct drift. Vision reports position, not bearing: a block on high points is geometrically correct but structurally deficient [1]. The cell therefore adds a seating check needing no extra hardware, since a force-limited arm senses contact: the height at which contact force develops is compared with that predicted from course datum and block dimension, a block seating high being lifted and re-placed.

A wall-plan sequencer converts the drawing into a course-by-course schedule. Because millimetric placement errors are especially detrimental in dry-stacked assemblies [15], each course is completed and verified before the next starts. It bounds the un-grouted lift height, since the wall is structurally incomplete until the mason grouts and reinforces the cells [21], so grouting is scheduled at intervals.

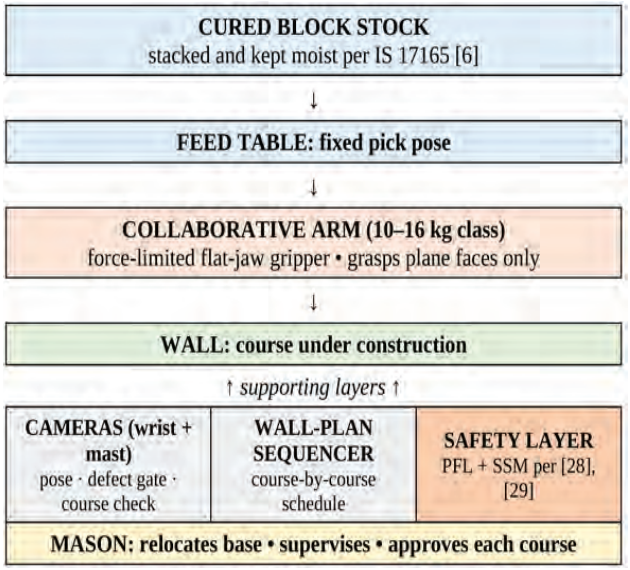


Fig. 2 Architecture of the collaborative ICEB laying cell

Operating Cycle

Fig. 3 gives the operating cycle, distinguishing the per-block loop (pick, gate, move, place and seat) from the per-course step of camera verification and mason approval. Rejected blocks leave at the gate, so the cell doubles as a quality filter on production.

Task Allocation

Table 2 states the human–robot split: the arm takes the repetitive, musculoskeletally punishing infill runs while the mason keeps every judgment task, including the reinforcement and grouting that give ICEB walls their ductile behaviour [21]. This is the division acceptance research envisions for construction robotics [27].

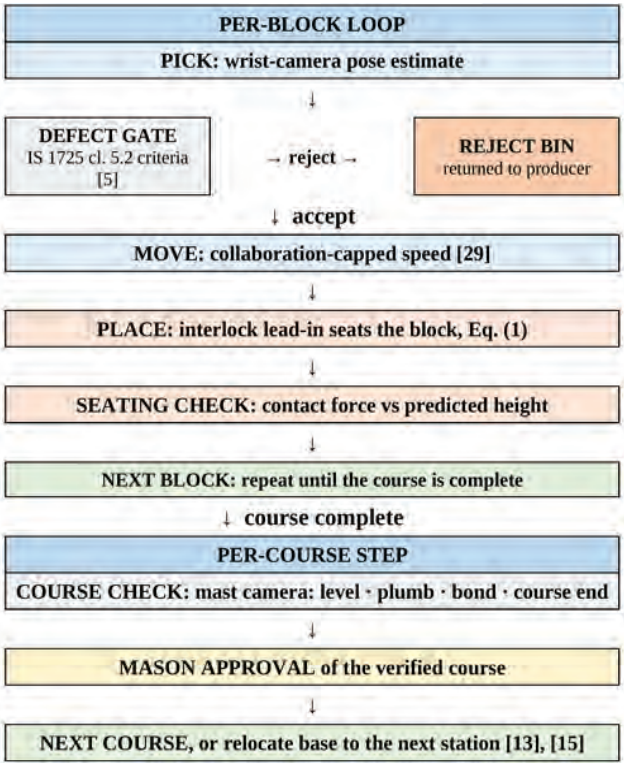


Fig. 3 Operating cycle: the per-block loop and the per-course verification step with mason approval

Table 2 Task allocation in the collaborative laying cell

Collaborative arm	Mason
Straight infill course laying	Setting out; first course on levelled bed
Per-block pick, gate, place	Corners, junctions, jambs and openings

Course geometry and seating checks	Reinforcement and cell grouting at scheduled intervals [21]
Repetitive lifting (up to 9.9 kg × hundreds)	Course approval; base relocation [13]
Halt on gate or check failure	Services, lintels, quality decisions

Safety as a Collaborative Application

The cell is designed as a collaborative application, the term used by ISO 10218-1:2025, which drops "collaborative robot" because only the application can be verified as collaborative [28]. The shared workspace uses power and force limiting with speed and separation monitoring at course approval, applying the biomechanical limits of ISO/TS 15066:2016, now in the ISO 10218:2025 series [29]: quasi-static limits of 140 N for hands and fingers, 160 N for the lower arm and wrist, 150 N for the upper arm and elbow. Hazards follow the construction-robotics taxonomy of physical risks, attentional costs and psychological impacts [25], met respectively by the force-limited grasp and capped speed, the pause at course approval, and the task split [27]. Grasp contact stress is capped below the IS 1725 strength class [5] per geometry-corrected protocols [24].

RESULTS AND DISCUSSION

Tolerance Budget

Table 3 evaluates Eq. (1) with worst-case inputs: the poorer demonstrated wall deviation, e_k = 2.90 mm [13], and the full codal tolerance t_b = ±2 mm [5]. The requirement is $c_{li} \geq 6.9$ mm, about 3.5 mm per side, a mould-design parameter rather than an exotic demand, since IS 1725 clause 7.1 permits shapes suited to construction needs through special mould attachments [5]. The binding term is cell-level registration across relocations, not arm repeatability, negligible at the ±0.05–0.1 mm typically specified for this class.

Course accumulation is the more demanding constraint, and the systematic term governs. Decomposing the ±2 mm envelope (t_{sys} = 0.5 mm with t_{ran} = 1.5 mm, or t_{sys} = 1.0 mm with t_{ran} = 1.0 mm), Eq. (2) gives end-of-course deviations of about 10–13 mm over ten blocks and 17–25 mm over twenty, three to seven times the 3.5 mm per-side clearance and beyond what any practical clearance absorbs uncorrected. Decisively, systematic bias is measurable: one calibration measurement per production batch establishes the effective block dimension, which the sequencer substitutes for the nominal value, removing the linear term and leaving a random residual of about 3–7 mm. The mast camera re-measures the realised course end each course, and the design reserves a closure block trimmed on site.

A trade-off must be acknowledged: the dry-joint shear capacity of [20] arises because the keys engage and break rather than slip, and widening the clearance to 3.5 mm per side introduces play in that load path, so whether the 0.035 MPa initial shear strength survives cannot be settled analytically. The clearance is a requirement to be validated for seating reliability and joint shear together.

Table 3 Tolerance budget for interlock seating, Eq. (1)

Quantity	Value	Basis
Cell placement error	2.90 mm	worst measured wall deviation [13]
Block dimensional tolerance	±2 mm	IS 1725:2023 cl. 7.3 [5]
Required lead-in clearance	≥ 6.9 mm total (≈3.5 mm/side)	Eq. (1), worst case
Design route	chamfered key–recess lead-in formed in mould	IS 1725 cl. 7.1, 7.4 [5]; [1], [14]

Unit Mass and Payload

Table 4 computes Eq. (3) between the codal minimum density of 1750 kg/m³ and a practical 1900 kg/m³, for solid units, hollow blocks [23] being lighter and the case therefore conservative. Computed as solid, the five codal modular sizes span 1.2 to 9.9 kg, the common 290 × 140 × 90 mm block giving 6.4–6.9 kg, which matches the measured 6.4–6.8 kg of production ICEBs [22]. A 2–3 kg gripper leaves 7–14 kg net across the 10–16 kg class, so a 12 kg arm carries the common unit at about 23 percent margin, while the heaviest modular size, 240 × 240 × 90 mm at up to 9.9 kg, requires a 16 kg arm, where the margin is 24 percent. Ordinary collaborative hardware therefore suffices across the codal range, against the 17 kg blocks of [11] and 50 kg units of [9] that pushed earlier systems to heavy machinery.

Table 4 Code-derived unit mass versus collaborative-arm payload

Block size (mm)	Volume (L)	Mass at 1750–1900 kg/m³ (kg)
290 × 90 × 90 [5]	2.35	4.1 – 4.5
290 × 140 × 90 [5]	3.65	6.4 – 6.9 (measured 6.4–6.8 [22])
305 × 143 × 100 [5]	4.36	7.6 – 8.3
240 × 240 × 90 [5]	5.18	9.1 – 9.9
Net arm capacity after a 2–3 kg gripper: 7–14 kg across the 10–16 kg class		

Cycle Time and Productivity

Table 5 assembles Eqs. (4)–(5) from conservative component times at capped speeds, giving 90–164 blocks per hour steady-state. Relocation must be added: Eq. (6) accommodates six to nine blocks per course and [13] averaged 15–20 blocks per relocation, so at an assumed two to three minutes each this amortises to 6–12 s per block, giving about 70–130 per hour effective, a band containing the realised 105 of [13]. The press rate of 30 blocks per hour in Table 5 is production, not laying, exposing the supply constraint, one cell consuming two to four presses; the mason figure of 400–500 units per day is fired brick [2] and context only.

Table 5 Cycle-time build-up and productivity comparison

Component / rate	Value	Basis
Pick + pose estimate	6–10 s	fixed pick pose, Fig. 2
Visual defect gate	4–8 s	camera criteria [5], [23]
Move to wall	8–14 s	capped collaborative speed [29]
Place + verify	4–8 s	interlock seating, Eq. (1)
Cycle → steady-state rate	22–40 s → 90–164 h^{-1}	Eqs. (4), (5)
Relocation amortised	6–12 s/block → 70–130 h^{-1}	15–20 blocks/ relocation [13]
Realised cobot rate	≈105 h^{-1}	138 blocks / ≈78 min incl. relocations [13]
Manual press output (production, not laying)	30 h^{-1} (3 workers)	[22]
Mason daily laying (fired brick; context only)	400–500 units	[2]

Workspace and Deployment

With reach 0.9–1.3 m and stand-off 0.35 m, Eq. (6) gives a usable frontage of 1.66–2.50 m per station, so a 10 m wall needs four to six stations, and courses above about 1.5 m need a lift axis, as platform geometry limited wall height in [13]. Economically the cell augments rather than replaces the crew, since Table 2 retains the mason and the feed table needs an attendant, confining the saving to the infill fraction. Near-term justification therefore rests on ergonomics, dimensional consistency and the documented shortage of construction workers [2], [27] rather than cost. The same trade is identified for robotic earthen construction, where operational and formwork savings are offset by capital and maintenance costs and precise quantification of market viability is left to further research, even though that route proved environmentally favourable [19]. Costing is deferred to the demonstrator.

How the Present Work Differs

Relative to prior investigators the distinction is efficiency through the unit rather than the machine. Commercial systems either retain a human to finish the mortar or deliver it under pressure through a purpose-built laying head [2]; thin-joint robots purchase tolerance with 1 mm rectified blocks [9], [11]; interlocking-assembly research purchases robot-friendliness with bespoke units [14]; wet earthen paths purchase automation with rheology control and in-place compaction [16], [17], [18]. This framework purchases none of these, adopting a codified unit whose ±2 mm tolerance, several-kilogram mass and load-bearing interlock are already standard [5], [20], [22]. Because that unit also showed the lowest lifecycle energy demand of the five systems compared

[3], and robotising earthen fabrication improves rather than erodes its environmental case [19], productivity and sustainability compound.

Implications for Skills and Technical Education

The reallocation in Table 2 changes what a mason does rather than displacing the mason, which carries a training requirement this journal's readership is placed to address. The retained tasks are the higher-skill portion of the trade; the transferred tasks are the repetitive lifting that causes musculoskeletal injury. Operating the cell adds competencies the masonry curriculum does not cover: reading a course plan against a digital wall model, performing the per-batch calibration on which Eq. (2) shows the system to depend, responding to gate and seating alarms, and working safely within a collaborative application [28], [29]. These suit a short add-on module in an ITI or diploma programme.

Limitations

The framework is analytical: no prototype has been built, and the evidence base leans heavily on one online-first source [13] whose demonstration used wooden blocks, so replicating it on production ICEBs is the first demonstrator task. The gripper is unproven against compressed-earth edge fragility, the mitigation being a flat-face force-limited grasp specified against geometry-corrected strength [24] with the gate rejecting damaged blocks [5]. Vertical reach beyond 1.5 m needs a lift axis, the feed table consumes labour, and the gain is bounded by the infill fraction. Durability against water, the reason IS 1725 caps absorption at 18 percent [5], remains a design responsibility, and adoption depends on the safety perceptions [25] and acceptance factors [27] documented for construction robotics.

A further caution concerns the codal tolerance itself: ±2 mm is a specification, not a measurement of what field presses deliver, and CEB production variability is appreciable, at a 24 percent coefficient of variation on unit strength [20]. The framework does not assume compliance so much as enforce it, since the gate rejects non-conforming blocks; but a cell fed by an uncontrolled press would reject heavily, making the rejection rate the decisive operating variable this analysis cannot predict and a demonstrator must measure.

CONCLUSIONS

The ICEB, conceived so that unskilled builders could dry-stack accurate walls, proves to be the masonry unit whose properties best match the documented failure modes of masonry robotics: every recurring barrier of bricklaying automation is removed by a property the unit already possesses. A lead-in clearance of 6.9 mm absorbs the codal tolerance and demonstrated cell accuracy at block level as ordinary mould design, but course-level accumulation is dominated by systematic mould bias and is tractable only with per-batch calibration, which the analysis establishes as a precondition. Code-derived masses suit ordinary collaborative arms, and an effective rate of 70–130 blocks per hour makes block supply rather than arm speed the binding constraint. Future work is a demonstrator: an instrumented mould realising the lead-in, grasp-and-place trials on production blocks, and a pilot wall with the gate and seating check in the loop. It confirms feasibility if placement error stays within the derived clearance over a course, the seating check detects nine in ten mis-seated blocks, rejection stays below one in ten, and the rate reaches 70 blocks per hour.

REFERENCES

1. V. Loing, O. Baverel, J.-F. Caron and R. Mesnil (2020), "Free-Form Structures from Topologically Interlocking Masonries", Automation in Construction, Vol. 113, Art. 103117. DOI: 10.1016/j.autcon.2020.103117
2. R. Bogue (2018), "What are the Prospects for Robots in the Construction Industry?", Industrial Robot, Vol. 45, No. 1, Pp. 1–6. DOI: 10.1108/IR-11-2017-0194
3. J. K. Shrestha (2021), "Assessment of Energy Demand and Greenhouse Gas Emissions in Low Rise Building Systems: Case Study of Five Building Systems Built after the Gorkha Earthquake in Nepal", Journal of Building Engineering, Vol. 34, Art. 101831. DOI: 10.1016/j.jobe.2020.101831
4. J. Bredenoord and Y. Kulshreshtha (2023), "Compressed Stabilized Earthen Blocks and Their Use in Low-Cost Social Housing", Sustainability, Vol. 15, No. 6, Art. 5295. DOI: 10.3390/su15065295
5. IS 1725:2023, Stabilized Soil Blocks Used in General Building Construction – Specification (Third Revision), Bureau of Indian Standards, New Delhi, 2023.
6. IS 17165:2020, Manufacture of Stabilized Soil Blocks – Guidelines, Bureau of Indian Standards, New Delhi, 2020.

7. K. Keyrouz, Y. Chen and F. Hamzeh (2025), "Integrating Modular Construction, Interlocking Blocks, and Robotic Technology with Lean Principles to Improve Masonry Construction Efficiency", Buildings, Vol. 15, No. 12, Art. 2133. DOI: 10.3390/buildings15122133

8. M. Schweiker, E. Endres, J. Gosslar et al. (2021), "Ten Questions Concerning the Potential of Digital Production and New Technologies for Contemporary Earthen Constructions", Building and Environment, Vol. 206, Art. 108240. DOI: 10.1016/j.buildenv.2021.108240

9. G. Pritschow, M. Dalacker, J. Kurz and M. Gaenssle (1996), "Technological Aspects in the Development of a Mobile Bricklaying Robot", Automation in Construction, Vol. 5, No. 1, Pp. 3–13. DOI: 10.1016/0926-5805(95)00015-1

10. S.-N. Yu, B.-G. Ryu, S.-J. Lim et al. (2009), "Feasibility Verification of Brick-Laying Robot Using Manipulation Trajectory and the Laying Pattern Optimization", Automation in Construction, Vol. 18, No. 5, Pp. 644–655. DOI: 10.1016/j.autcon.2008.12.008

11. Z. Dakhli and Z. Lafhaj (2017), "Robotic Mechanical Design for Brick-Laying Automation", Cogent Engineering, Vol. 4, No. 1, Art. 1361600. DOI: 10.1080/23311916.2017.1361600

12. E. P. G. Bruun, R. Pastrana, V. Paris et al. (2021), "Three Cooperative Robotic Fabrication Methods for the Scaffold-Free Construction of a Masonry Arch", Automation in Construction, Vol. 129, Art. 103803. DOI: 10.1016/j.autcon.2021.103803

13. W. W. Loy, M. Belek Fialho Teixeira and A. P. Franzè (2026), "AR-Guided Modular Robotic Bricklaying: A Proof-of-Concept System for Human-in-the-Loop Construction Workflows", Construction Innovation, online-first. DOI: 10.1108/CI-07-2025-0341

14. Y. Zhang, Y. Koga and D. Balkcom (2021), "Interlocking Block Assembly with Robots", IEEE Transactions on Automation Science and Engineering, Vol. 18, No. 3, Pp. 902–916. DOI: 10.1109/TASE.2021.3069742

15. Y. Liu and N. Napp (2023), "Stability-Based Sequence Planning for Robotic Dry-Stacking of Natural Stones", IEEE Robotics and Automation Letters, Vol. 8, No. 9, Pp. 5894–5901. DOI: 10.1109/LRA.2023.3301275

16. A. Perrot, D. Rangeard and E. Courteille (2018), "3D Printing of Earth-Based Materials: Processing Aspects", Construction and Building Materials, Vol. 172, Pp. 670–676. DOI: 10.1016/j.conbuildmat.2018.04.017

17. M. Gomaa, W. Jabi, A. Veliz Reyes and V. Soebarto (2021), "3D Printing System for Earth-Based Construction: Case Study of Cob", Automation in Construction, Vol. 124, Art. 103577. DOI: 10.1016/j.autcon.2021.103577

18. L. P. Schmitz, J. Gosslar, E. Dorresteijn et al. (2024), "Experimental Investigations on the Compaction Energy for a Robotic Rammed Earth Process", Frontiers in Built Environment, Vol. 10, Art. 1363804. DOI: 10.3389/fbuil.2024.1363804

19. M. Lange, J. Gosslar, S. V. Albrecht et al. (2026), "Comparative Life Cycle Assessment of Manual and Robotic Fabrication of an Unstabilized Rammed Earth Wall", Buildings, Vol. 16, No. 10, Art. 1897. DOI: 10.3390/buildings16101897

20. T. Sturm, L. F. Ramos and P. B. Lourenço (2015), "Characterization of Dry-Stack Interlocking Compressed Earth Blocks", Materials and Structures, Vol. 48, No. 9, Pp. 3059–3074. DOI: 10.1617/s11527-014-0379-3

21. B. Qu, B. J. Stirling, D. C. Jansen et al. (2015), "Testing of Flexure-Dominated Interlocking Compressed Earth Block Walls", Construction and Building Materials, Vol. 83, Pp. 34–43. DOI: 10.1016/j.conbuildmat.2015.02.080

22. M. Sassu, A. Romanazzi, L. Giresini et al. (2018), "Production Procedures and Mechanical Behaviour of Interlocking Stabilized Compressed Earth Blocks (ISCEBs) Manufactured Using Float Ram 1.0 Press", Engineering Solid Mechanics, Vol. 6, No. 2, Pp. 89–104. DOI: 10.5267/j.esm.2018.3.004

23. P. Kasinikota and D. D. Tripura (2022), "Prediction of Physical-Mechanical Properties of Hollow Interlocking Compressed Unstabilized and Stabilized Earth Blocks at Different Moisture Conditions Using Ultrasonic Pulse Velocity", Journal of Building Engineering, Vol. 48, Art. 103961. DOI: 10.1016/j.jobe.2021.103961

24. J.-C. Morel, A. Pkla and P. Walker (2007), "Compressive Strength Testing of Compressed Earth Blocks", Construction and Building Materials, Vol. 21, No. 2, Pp. 303–309. DOI: 10.1016/j.conbuildmat.2005.08.021

25. Y. Sun, I. Jeelani and M. Gheisari (2023), "Safe Human-Robot Collaboration in Construction: A Conceptual Perspective", Journal of Safety Research, Vol. 86, Pp. 39–51. DOI: 10.1016/j.jsr.2023.06.006

26. C.-J. Liang, K. M. Lundeen, W. McGee et al. (2019), "A Vision-Based Marker-Less Pose Estimation System for Articulated Construction Robots", Automation in Construction, Vol. 104, Pp. 80–94. DOI: 10.1016/j.autcon.2019.04.004

27. S. Park, H. Yu, C. C. Menassa and V. R. Kamat (2023), "A Comprehensive Evaluation of Factors Influencing Acceptance of Robotic Assistants in Field Construction Work", Journal of Management in Engineering, Vol. 39, No. 3, Art. 04023010. DOI: 10.1061/JMENEA.MEENG-5227

28. ISO 10218-1:2025, Robotics – Safety Requirements – Part 1: Industrial Robots (Third Edition), International Organization for Standardization, Geneva, 2025.

29. ISO/TS 15066:2016, Robots and Robotic Devices – Collaborative Robots, International Organization for Standardization, Geneva, 2016.

DOI - 10.68648/ijte-q-2026-49-03-005

Similitude Framework for Scale Modelling of Dry-Stacked Interlocking Compressed Earth Block Walls

Vishal Kumar Chaudhary
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Khagaria
Khagaria
✉ vishal@geckhagaria.org.in

Sneha Kumari
Assistant Professor
Dept. of Civil Engineering
Gaya College of Engineering
Gaya
✉ sneha21.dstte@bihar.gov.in

Sumit Suman
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Khagaria
Khagaria
✉ suman.sumit@geckhagaria.org.in

ABSTRACT

Interlocking compressed earth blocks (ICEBs) are machine-pressed soil blocks that stack together like toy bricks, without mortar. They offer a low-cost, low-carbon way to build houses in earthquake-prone regions. Laboratory testing of ICEB walls remains rare because full-size tests need large and costly equipment. Testing small models instead would solve this problem, but only if clear rules exist to make the small model behave like the real wall. Such rules, called similitude rules, do not yet exist for dry-stacked interlocking earthen walls, because these walls act through friction and interlocking between blocks rather than as one solid material. This paper derives the missing rules using dimensional analysis. Three ratios must be matched between model and prototype: the friction between block faces, the ratio of key strength to friction resistance, and the ratio of vertical load to block strength. The paper shows how to correct the reduced self-weight of a small model with added weights, how to correct for size effects, and how to press model blocks in 3D-printed moulds. A quarter-scale worked example based on published test data requires only a 37 kg added weight and one sixteenth of the lateral testing force.

***KEYWORDS** : Compressed earth block, Dimensional analysis, Dry-stack masonry, Scale modelling, Similitude.*

NOTATIONS

λ: scale factor (model size divided by prototype size)
L: block length
H: wall height
f_b: block compressive strength
f'm: masonry (prism) strength
E: modulus of elasticity
ρ: density
g: gravitational acceleration
μ: friction coefficient between bed faces
τ_k: key shear strength
A_k, A_c: key area and block contact area
σ_v: vertical stress at a joint
σ_n: normal stress at the joint considered
σ_r: roof (service) surcharge
$\Delta\sigma$: added compensating surcharge
$\hat{g}$: joint gap
F: lateral force
δ: displacement
a: acceleration
t: time (dynamic loading)
t_w: wall thickness

κ_f, κ_k: measured size-effect corrections on block and key strength

Π_1, Π_2, Π_3: the three governing dimensionless ratios; subscripts m, p: model, prototype.

INTRODUCTION

An interlocking compressed earth block (ICEB) is made by pressing local soil, mixed with a small amount of cement (usually 5 to 10 per cent), in a manual press. The block comes out with raised keys on top and matching recesses below, together with vertical holes, so that blocks stack onto each other like toy bricks. No mortar is needed, and steel bars with grout are added only where the design requires them [1]. Because the main raw material is soil from near the site, ICEB houses are inexpensive, need little transport, save energy compared with fired bricks or concrete, and can be built quickly by workers with basic training. Recent projects have even used computer-customized ICEB shapes for housing in low-resource regions [2]. For India, where affordable rural housing must often be built in earthquake-prone areas, this technology is directly relevant.

Almost everything known about how ICEB walls resist earthquake-type loading comes from a small number of full-size laboratory tests: in-plane cyclic tests on shear walls [3], out-of-plane tests on reinforced walls [4], tests on flexure-dominated walls [5], [6], a study of how block shape affects wall performance [7], a seismic design manual built on these results [8], and one full-size dry-stack building tested in South Africa under experimentally applied earthquake loading [9]. Every one of these programmes needed a strong floor, a heavy reaction frame, hundreds to thousands of blocks, and months of work. Very few institutions in the regions where ICEB houses are actually built have such facilities.

The usual engineering answer to this problem is to test a small-scale model instead of the real structure, the same idea as testing a model aircraft in a wind tunnel. This is already being attempted for earthen blocks: reduced-scale single-storey block building models have been tested on a shaking table in India [10]. The difficulty is that these tests reduced only the geometry. They did not follow any rules for scaling the weight, the friction, the key strength or the roof load, so their results cannot be converted back to full size with confidence. The only published study that scaled an interlocking earth-block structure rigorously used a geotechnical centrifuge, spinning additively manufactured model blocks at elevated acceleration [11]. That approach works, but a centrifuge is a rare national facility, so the accessibility problem remains.

This paper fills the gap. It derives, from first principles, the complete set of rules (the similitude framework) needed to test small models of dry-stacked ICEB walls in an ordinary laboratory. It identifies the three ratios that must be matched between model and prototype, shows how to correct the one unavoidable error (the reduced self-weight of the model) with simple added weights, explains how to handle size effects, and gives a practical recipe for making model blocks in 3D-printed moulds. A worked example at one-quarter scale, using only published full-scale data, shows that everything needed fits on a laboratory bench.

LITERATURE REVIEW

Three bodies of research meet in this work: how dry-stacked interlocking walls behave; how scale models are designed for other structures; and how strength changes with specimen size. A systematic Scopus search (ten queries, August 2026, combining interlocking, compressed earth block, dry-stack and mortarless terms with wall, shear and seismic terms, similitude, scale model and dimensional analysis terms, size-effect terms, and 3D-printed mould terms) returned 1,850 records, of which 1,783 were unique after removing duplicates. The sources below were all read in full text.

How dry-stacked interlocking walls behave

Laboratory characterization of dry-stack ICEBs shows that wall strength depends on four factors: the stabilized soil itself, the friction between block faces, the contact at the interlocking keys, and the geometry. Joint shear follows the Coulomb friction law, and the initial shear resistance comes from the interlock indentations rather than from any bond; in shear tests at low compression the indentations broke [1]. Measured peak friction coefficients on dry compressed earth block interfaces are high: values of about 1.20 and 1.07 were obtained for two stabilized mixtures, far above the value of 0.4 that Eurocode 6 proposes for mortared joints [12]. Machine-pressed stabilised soil blocks used for

structural applications in India are governed by IS 1725:2023, which specifies dimensions , compressive strength classes and physical requirements; no BIS provision yet addresses the design or seismic testing of dry-stacked interlocking walls, and the evidence that scale modelling generates is precisely what a future Indian provision would require. The vertical load a wall can carry depends on its slenderness and on how well the blocks actually touch [13], and the out-of-plane strength of dry-stack walls rises directly with the vertical load pressing the blocks together [14]. The same joint-controlled behaviour is seen in mortarless interlocking walls of other materials: such walls dissipate earthquake energy mainly by blocks sliding on each other [15], and a recent numerical simulation of an ICEB shear wall with dry vertical and horizontal joints confirms that the joint parameters control the response [16]. Two joint-level details matter especially. First, block faces are never perfectly flat, and this unevenness concentrates the contact and changes the capacity, so flatness is a real physical variable [17]. Second, cyclic friction tests on dry mortarless surfaces show that shear resistance varies linearly with normal stress, following the Mohr-Coulomb criterion without cohesion; within the tested range the friction coefficient is therefore constant [18]. This last fact, as the Methodology shows, is what makes small-scale testing of these walls possible at all.

The full-scale ICEB tests [3] to [8] also reveal something unusual: standard masonry design formulas transfer poorly to this system, and in different ways for different mechanisms. For the shear-dominated walls, the ACI 530 method significantly overestimated the measured in-plane shear capacity [3]. For the out-of-plane walls, the flexural strength was predicted adequately [6], but the stiffness was not: measured displacements were several times larger than the code predictions, the measured masonry modulus was 350 MPa (equal to 125 times the masonry strength, against the ACI 530 approximation of 700 times for clay masonry), the strain at peak prism strength was about 0.014 against typical code values of 0.0025 to 0.003, and a joint gap of 0.4 mm per course had to be introduced in the analysis to match the measured response [4]. In plain terms, these walls move by blocks rotating and sliding at the joints, not by the material stretching. That is exactly the behaviour that ordinary scaling rules, written for solid materials, cannot handle.

Scale-model theory and practice

The mathematics behind scale models is dimensional analysis, formalized by the Pi theorem: any physical relationship can be rewritten in terms of a small number of dimensionless ratios, and if a model matches the real structure in those ratios, it behaves in the same way [19]. Structural engineers have a standard handbook of such rules covering static and dynamic models, model materials, the well-known gravity problem of small models, and size effects [20]. For ordinary mortared masonry the practice is mature: a reinforced masonry building has been tested on a shaking table at a length scale of 1:3, selected from the dimensional and load capacities of the table [21]; methods exist to compensate small unavoidable rule violations, called distortions [22]; and researchers actively compare the practical modelling options of true replica, added mass, and accepted distortion [23]. All of this practice, however, assumes the wall behaves like a continuous material. No such rulebook exists for walls that act through dry friction and interlocking keys.

Size effects

Small specimens of brittle materials are, size for size, stronger than large ones. For masonry in compression this is well documented through prism and wallette testing [24]. Even the drift a wall can sustain before failing decreases as the wall gets larger [25]. For earthen walls specifically, testing at several scales showed that small specimens gave a cohesion of 0.0715 MPa while a full-scale two-storey wall (5.95 m by 6.20 m by 0.65 m) of the same material gave a much lower cohesion of 0.0139 MPa; small-scale specimens consistently overestimate shear strength [26]. In India, machine-pressed stabilised soil blocks for structural use are governed by IS 1725:2023 [27], which sets dimensional, strength and durability requirements for the unit but not for wall assemblies. Seismic guidance for earthen buildings is given separately in IS 13827:1993 [28], whose scope is expressly limited to earthen houses built without stabilisers such as cement or lime, and which therefore does not extend to cement-stabilised interlocking blocks. No Indian provision presently covers the seismic design or the scale-model

testing of dry-stacked interlocking walls, and evidence of the kind this framework generates is precisely what such a provision would require. A small-model method for earthen walls must therefore include measured corrections for size, not ignore them.

The gap

Putting these strands together: small-scale earthen-block testing is already attempted, but without rules [10]; the rules exist, but only for solid, mortared masonry [20] to [23]; and the single study at the exact intersection of compressed earth blocks, dry-stacking and scale modelling (one paper out of 1,783 found in the systematic search) relies on a centrifuge [11]. To the authors' knowledge, no published framework tells an ordinary laboratory how to build and test a valid small-scale model of a dry-stacked interlocking earthen wall. This paper provides that framework.

METHODOLOGY

What must be matched: the three governing ratios

Consider a dry-stacked interlocking wall described by its block length L, with block height, block width, key dimensions, wall height H and wall length all held in fixed proportion to L as geometric ratios; block strength f_b; modulus E; density ρ; friction coefficient μ between bed faces; key shear strength τ_k acting over key area A_k on block contact area A_c; vertical stress σ_v from self-weight and roof load; gravity g; joint gap $\hat{g}$; an applied lateral force F; the resulting movement δ; and, for dynamic loading, acceleration a and time t.

Table 1. Variables of the dry-stacked interlocking wall problem and their dimensions

Variable	Symbol	Dimension (M-L-T)
Block length (reference length)	L	L
Wall height; joint gap; displacement	H; $\hat{g}$; δ	L
Block strength; modulus; key shear strength; vertical stress	f_b; E; τ_k; σ_v	$M L^{-1} T^{-2}$
Density	ρ	$M L^{-3}$
Gravity; acceleration	g; a	$L T^{-2}$
Time	t	T
Lateral force	F	$M L T^{-2}$
Friction coefficient	μ	1 (dimensionless)

Table 1 lists the fourteen variables. Their dimensional matrix has rank three (M, L, T), so the Pi theorem [19] gives 14 − 3 = 11 independent dimensionless groups: the three geometric ratios H/L, $\hat{g}$/L and δ/L; the friction coefficient μ, which is dimensionless and enters the set directly; and seven mechanical groups, E/f_b, τ_k/f_b, σ_v/f_b, $\rho gL/f_b$, a/g, t^2g/L and $F/(f_bL^2)$. The physically meaningful ratios used below are formed from these by multiplication; in particular Π_2 combines τ_k/f_b, σ_n/f_b, μ and the geometric key-to-contact area ratio A_k/A_c. Three ratios govern the strength and failure mode of the wall:

$$\Pi_1 = \mu = \tau / \sigma_n \quad (1)$$

The friction between block faces. This decides when blocks start to slide on each other. In the same-material regime, the friction coefficient is matched automatically; it is also measured in calibration step (3).

$$\Pi_2 = \tau_k A_k / (\mu \sigma_n A_c) \quad (2)$$

Key strength compared with friction strength. This decides which happens first, sliding on the bed face or breaking of the keys, and therefore selects the failure mode. With the stress scale equal to one, this ratio is matched pointwise at every joint of the wall.

$$\Pi_3 = \sigma_v / f_b \quad (3)$$

Vertical load compared with block strength. This governs rocking, joint opening and crushing. It is the ratio the gravity distortion attacks.

The physical importance of these ratios is confirmed directly by experiments: dry joints obey the friction law with the keys supplying the initial shear resistance [1]; wall capacity rises with vertical pre-compression [14]; and axial capacity depends on the contact state [13]. Secondary ratios complete the set: the self-weight ratio $\rho gL/f_b$ (the source of the gravity problem, discussed below), the stiffness ratio E/f_b, the scaled joint gap $\hat{g}$/L (the analysis-derived full-scale gap is about 0.4 mm per course [4]), and, for dynamic testing, a/g and t^2g/L. If model and prototype share the same values of the three governing ratios, and the secondary ratios are respected or corrected, the model wall will reproduce the prototype

strength and failure mechanism at the properly scaled load, with deformation quantities interpreted subject to the size-effect corrections described below.

The one unavoidable error, and how to fix it

The simplest and most reliable way to build the model is from the same material: press the model blocks from the same stabilized soil as the real ones. Then the strength, the density and, because the sliding surfaces are the same material pair, the friction coefficient are all nominally the same, and Π_1 is matched. One qualification applies: because the model soil must be sieved to a scaled maximum particle size (see Experimental setup below), it is an approximation to the prototype material with a modified grading, and that is precisely why μ and the strength corrections are measured in the calibration sequence rather than assumed. With the same material, stresses in the model equal stresses in the prototype, so all forces simply scale with area: a model at scale λ = 1/4 needs only 1/16 of the testing force (Table 2).

One thing, however, does not scale: gravity. A quarter-scale wall is only a quarter as tall, so the squeeze that its own weight puts on the bottom joints is only a quarter of the real value. The model is too light for its strength. The centrifuge fixes this by spinning the model to raise the effective gravity [11]. Our framework, meant for ordinary laboratories, fixes it instead with added weights, following the established idea of artificial mass simulation [20] and explicit distortion compensation [22]. The rule is simple. Add, at the top of the model wall, a surcharge stress

$$\Delta\sigma = (1-\lambda)\cdot\rho g H + \sigma_r \qquad (4)$$

where H is the prototype wall height and σ_r is the roof load the real wall carries. This makes Π_3 exactly right at the bottom joint, where walls usually fail. The joints higher up then carry slightly more relative load than in the real wall. This error is bounded, is quantified in the worked example, and can be removed when necessary by splitting the surcharge into small weights placed every few courses. Most importantly, the correction is safe: because the friction coefficient is constant over the working stress range [18], adding weights to fix Π_3 does not disturb Π_1. That single experimental fact is what makes small-scale testing of these walls at normal gravity valid. For dynamic (shaking-table) testing one further requirement applies: inertial similitude at acceleration scale one requires distributed added mass equal to $(1/\lambda - 1)$ times the structural mass of the model, attached along the height [20]; this added mass simultaneously supplies the gravity-stress correction described above.

Size-effect corrections

Even with the same soil, small pressed blocks tend to come out somewhat stronger than large ones, and small specimens overestimate shear strength [24], [26]. The framework does not assume these effects away; it measures them. Two correction factors are determined by test: κ_f (model block strength divided by prototype block strength, from compression tests) and κ_k (the same ratio for key shear strength). Because key shear is a brittle fracture process, its size dependence follows fracture-type size-effect behaviour rather than simple strength proportionality; the measured κ_k therefore stands in for a size-effect law, cannot be assumed equal to one, and may itself depend on λ. All strengths measured on the model are divided by the applicable κ before being reported at full scale, and deformation results carry the additional caution that drift capacities themselves decrease with size [25].

Experimental setup: making and checking the model blocks

Model blocks are pressed in 3D-printed plastic moulds, an approach already proven for scaled interlocking blocks [11]. The mould is simply the block shape in negative, printed with three requirements: every key and hole scaled by exactly λ, so that the failure-mode ratio Π_2 is preserved (block shape is known to strongly affect wall behaviour [7]); small draft angles of 2 to 3 degrees so blocks release cleanly; and a printed-face flatness better than the scaled joint gap, $\hat{g}_m = \lambda \times 0.4$ mm, because face unevenness genuinely changes capacity [17]. Because the compaction pressure needed to reach the prototype dry density is far beyond what an unreinforced polymer face can carry repeatedly, the printed part serves as a geometry-defining insert seated in a steel or aluminium sleeve, with the pressing load carried by the metal frame; insert flatness is re-verified at regular pressing intervals and the mould service life is recorded and reported. Two tolerances are controlled separately, both scaled by λ: the face flatness of the mould, and the standard deviation of pressed block

heights, because course-height variability, not mould flatness alone, generates the joint gaps observed at full scale [4]. Curing is matched to prototype practice: the same mix, the same humid-cure regime and the same total curing duration are used [3], [4], and every friction and key-shear test is carried out at a controlled and reported moisture content, because the strength and the friction of earthen materials are moisture-sensitive. Fig. 1 shows the block scaling concept. Pressing is mass-controlled: each charge of soil is weighed to the scaled block mass and pressed to the scaled height, reproducing the real block density. The soil is sieved so that the largest particle also scales (the 4 mm full-scale pulverizer limit [4] becomes 1 mm at quarter scale), and the moisture is set by the standard drop test [4]. A minimum-scale criterion follows from the sieving rule: the smallest key dimension of the model, equal to λ times its prototype value, must span at least five maximum particle diameters for the key to behave as a continuum element rather than a particle-count-dominated one; this is the limit given for model materials in established practice, where the ratio of the largest particle size to the smallest model dimension is required not to exceed one fifth [20]. For brittle key features a stricter margin of up to ten diameters is prudent. With a 1 mm maximum particle at quarter scale, this requires model key features of at least 5 mm (10 mm with the stricter margin), corresponding to prototype key features of at least 20 mm (40 mm); block geometries with finer keys sit at or beyond this bound at $\lambda = 1/4$, and smaller scale factors are then inadmissible with this soil. This criterion defines a minimum admissible scale for a given block geometry and soil, and it is one more reason the key-shear calibration of step (4) is mandatory rather than optional: the measured κ_k absorbs both the size effect and any fabrication effect near the bound.

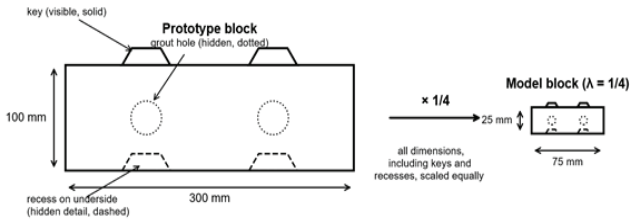


Fig. 1 Prototype and quarter-scale model block: every dimension, including the interlock keys, is multiplied by $\lambda = 1/4$. Key geometry is shown schematically; actual moulds follow the prototype block drawings.

Before any wall is built, five short calibration tests measure every quantity the framework needs, in the order shown in Fig. 2: (1) size, height-variability and flatness checks on the pressed blocks; (2) compression tests giving κ_f; (3) block-pair friction tests under a range of clamping pressures, analysed by the Mohr-Coulomb method of [18] and compared with the published full-scale values of about 1.07 to 1.20 [12]; (4) key shear tests giving τ_k and κ_k; and (5) prism (stacked block) compression giving the masonry strength and the joint-gap statistics (mean and coefficient of variation) for comparison with the full-scale value [4]. Minimum replication is prescribed: five blocks in compression, five friction pairs tested at no fewer than four normal-stress levels, five key-shear pairs and three prisms, with the mean and coefficient of variation reported for every parameter and the κ factors defined on means. Only when the three governing ratios are demonstrably matched does wall testing begin.

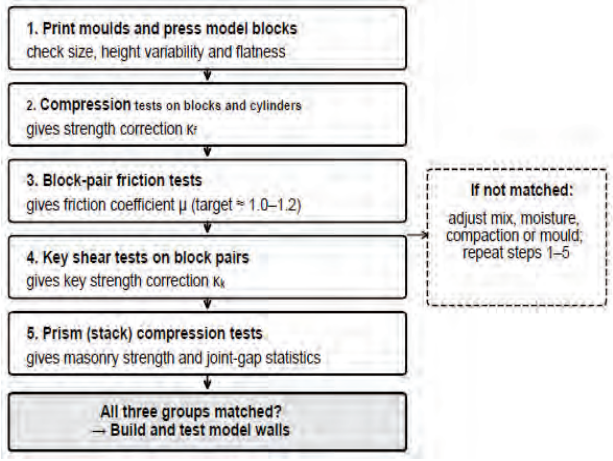


Fig. 2 The five-step calibration sequence: every contact parameter is measured before any wall is tested

RESULTS AND DISCUSSION

The scaling rules

Table 2 lists the complete set of conversion factors between model and prototype for same-material testing at ordinary gravity. Reading it practically: at quarter scale, lengths and movements shrink to one quarter, stresses and drifts stay the same, forces drop to one sixteenth, and specimen mass drops to one sixty-fourth. A wall test that needs a heavy reaction frame at full scale becomes a bench-top test.

Table 2. Conversion factors for same-material scale models of dry-stacked ICEB walls

Quantity	Scale factor (model/ prototype)	At $\lambda = 1/4$
Length, displacement	λ	0.25
Stress, strength, friction, drift	1	1
Force	λ^2	1/16
Mass	λ^3	1/64
Self-weight stress	λ (deficient; corrected by added weights)	0.25
Added surcharge	$\Delta\sigma = (1 - \lambda)\rho g H + \sigma_r$	see Table 3
Time / frequency (dynamic)	$\sqrt{\lambda}$ / $1/\sqrt{\lambda}$ (needs distributed added mass; see Methodology)	0.5 / 2.0

Worked example at quarter scale

The rules are now applied to the best-documented system in the literature: the 300 mm × 150 mm × 100 mm interlocking block with measured unit strength 7.76 MPa, grouted prism strength 2.81 MPa and modulus 350 MPa [4], with a minimum block dry density of 1850 kg/m³ (18.15 kN/m³) [3], adopted here as the block density, and a standard-block net-to-gross area ratio of 39,320/45,000 = 0.874 [4], built into an ungrouted dry-stacked wall of the same 1.8 m × 1.8 m × 0.15 m geometry as the shear walls of [3]. The example addresses the ungrouted configuration that the framework primarily targets; where grouted and reinforced walls are to be modelled, additional ratios enter (grout-to-block strength, reinforcement ratio and steel-yield-to-block-strength), the fully grouted density of 18 kN/m³ applies [4], and the scaling of reinforcement is a recognized further difficulty [20]. Fig. 3 illustrates the result and Table 3 gives the numbers. The model blocks measure 75 mm × 37.5 mm × 25 mm; the model wall is 450 mm square and 37.5 mm thick. The block density and the net-area ratio give an ungrouted wall unit weight of 0.874 × 18.15 = 15.9 kN/m³. The weight of the real wall therefore presses its base joints with 15.9 × 1.8 = 28.6 kPa; the model wall provides only 15.9 × 0.45 = 7.2 kPa; the difference of 21.4 kPa, acting on the model plan area of 0.45 m × 0.0375 m = 0.016875 m², requires a force of about 362 N, that is an added mass of about 37 kg (taking g = 9.81 m/s²), plus the scaled roof load where one exists. The soil is sieved to 1 mm, and mould flatness is held to 0.1 mm.

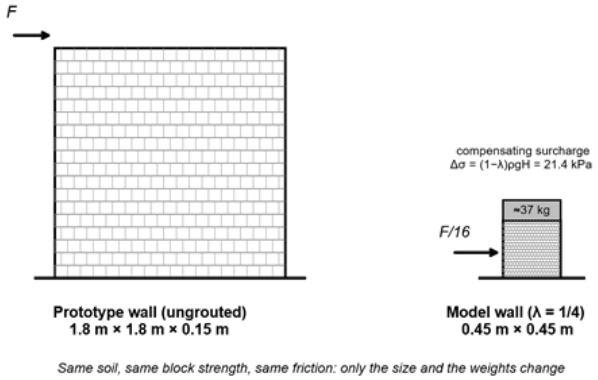


Fig. 3 The quarter-scale model of the reference wall: one sixteenth of the force, plus a 37 kg weight to restore the missing self-weight stress.

Table 3. Worked example: quarter-scale model built entirely from published full-scale data

Quantity	Prototype [3], [4]	Model ($\lambda = 1/4$)
Block (L × W × H)	300 × 150 × 100 mm	75.0 × 37.5 × 25.0 mm
Wall (L × H × t_w)	1800 × 1800 × 150 mm	450 × 450 × 37.5 mm
Block strength target	7.76 MPa	7.76 MPa (divided by measured κ_f)
Friction μ target	about 1.07 to 1.20 [12]	same (verified by test)
Joint gap / mould flatness	about 0.4 mm, analysis-derived [4]	0.1 mm or better
Largest soil particle	4 mm [4]	1 mm (sieved)
Wall unit weight (ungrouted)	15.9 kN/m³ (derived from [3], [4])	same material
Self-weight stress at base	28.6 kPa	7.2 kPa + 21.4 kPa added
Added weight on model	none	about 37 kg (+ scaled roof load)
Testing force	F	F/16

Three points deserve discussion. First, all three governing ratios can be satisfied at once, with entirely ordinary means: the same soil, a printed mould in a metal

sleeve, a bench press, and a 37 kg weight. The surcharge should be applied so that it follows the wall without restraining it, either as a gravity mass seated through a low-friction bearing or as a servo-controlled vertical force; with a gravity mass, the P-delta contribution of the concentrated weight approximates, but does not exactly reproduce, that of the distributed self-weight of the prototype, and the application method must be reported. Second, the one remaining error is known exactly. With a single top weight, the upper joints of the model carry up to 21.4 kPa more than strict similarity requires. Against a block strength of 7.76 MPa this is negligible (an error in Π_3 of about 0.003), but at very low clamping pressures it could slightly delay sliding at the top courses. If the expected failure involves those courses, for example rocking or out-of-plane bending, the weight should be split into smaller masses placed every few courses, exactly as the Methodology prescribes. Third, the mould-flatness target of 0.1 mm is the hardest requirement in the whole framework, and it must be treated as an acceptance check on every mould, because face unevenness genuinely changes how these walls carry load [17].

Comparison with other approaches, scope and limitations

Compared with the centrifuge route [11], the present framework trades perfect gravity scaling for accessibility, and replaces it with a correction whose size is known. Centrifuge results are in fact the ideal benchmark against which walls built under this framework should first be validated. Compared with geometry-only scaling [10], the framework adds precisely what was missing, namely the weight rule, the friction and key checks, and the size corrections, so that model results become transferable to full size. Because the three ratios come from joint mechanics shared by all mortarless interlocking walls, confirmed in interlocking brick [15] and reproduced numerically for ICEB walls [16], the framework extends, with re-calibration, to dry-stack interlocking masonry in general. Its limits are stated plainly. It covers quasi-static and moderate-rate loading; the dynamic extension by distributed added mass follows [20], [23] but is not demonstrated here. Small pressed specimens dry differently from full-size blocks, so moisture must be controlled and reported. The corrections κ_f and κ_k, and the caution that drift capacities decrease with size [25], [26], must always accompany any transfer of results. Equally, the framework guarantees similitude of strength and failure mechanism; stiffness and displacement quantities depend additionally on the joint-gap distribution, which the flatness and height-variability controls bound from above but do not reproduce statistically, so model deformations are transferred only together with the measured prism joint-gap statistics of step (5). Finally, the framework is proposed and analytically justified; its experimental validation is the natural next step, for which the calibration sequence of Fig. 2 is the ready-made programme.

CONCLUSIONS

This paper has derived the first set of rules for testing small-scale models of dry-stacked interlocking compressed earth block walls in an ordinary laboratory. The key findings are: (i) three dimensionless ratios govern these walls, namely the friction between block faces, the key-strength-to-friction ratio, and the vertical-load-to-block-strength ratio, and a model matching all three reproduces the strength and failure mechanism of the real wall; (ii) building the model from the same soil matches the friction ratio automatically, and the single unavoidable error, the deficient self-weight of the model, is corrected by an added weight computed from a one-line formula, a correction that does not disturb the friction ratio because friction in these joints is constant over the working stress range; (iii) size effects are handled by two measured correction factors built into a five-step calibration sequence, rather than by assumption; (iv) a quarter-scale worked example built on published data needs only a 37 kg weight, 1 mm sieved soil, 0.1 mm mould flatness and one sixteenth of the lateral testing force, placing meaningful ICEB wall research within reach of any institution with a bench press and a desktop 3D printer; and (v) 3D-printed moulds with weighed, height-controlled pressing give model blocks of faithful geometry at negligible cost. The novelty lies in extending scale-model theory from solid, mortared masonry to walls governed by friction and interlocking. Future work should validate the framework experimentally through the proposed calibration and wall tests, benchmark it against the published centrifuge results, and extend it to shaking-table testing using distributed added masses.

REFERENCES

1. Sturm T., Ramos L.F. and Lourenço P.B. (2015), "Characterization of Dry-Stack Interlocking Compressed Earth Blocks", Materials and Structures, Vol. 48, Pp. 3059–3074. DOI: https://doi.org/10.1617/s11527-014-0379-3.
2. Zhang Y., Tatarintseva L., Clewlow T., Clark E., Botsford G. and Shea K. (2023), "A Template Design and Automated Parametric Model for Sustainable Corbel Dwellings with Interlocking Blocks", Developments in the Built Environment, Vol. 14, Art. 100148. DOI: https://doi.org/10.1016/j.dibe.2023.100148.
3. Bland D.W. (2011), In-Plane Cyclic Shear Performance of Interlocking Compressed Earth Block Walls, M.S. Thesis, California Polytechnic State University, San Luis Obispo. DOI: https://doi.org/10.15368/theses.2011.61.
4. Herskedal N.A. (2012), Investigation of Out-of-Plane Properties of Interlocking Compressed Earth Block Walls, M.S. Thesis, California Polytechnic State University, San Luis Obispo. DOI: https://doi.org/10.15368/theses.2012.235.
5. Qu B., Stirling B.J., Jansen D.C., Bland D.W. and Laursen P.T. (2015), "Testing of Flexure-Dominated Interlocking Compressed Earth Block Walls", Construction and Building Materials, Vol. 83, Pp. 34–43. DOI: https://doi.org/10.1016/j.conbuildmat.2015.02.080.
6. Laursen P.T., Herskedal N.A., Jansen D.C. and Qu B. (2015), "Out-of-Plane Structural Response of Interlocking Compressed Earth Block Walls", Materials and Structures, Vol. 48, Pp. 321–336. DOI: https://doi.org/10.1617/s11527-013-0186-2.
7. Ambers S.E. (2017), In-Plane Shear Wall Performance as Affected by Compressed Earth Block Shape, M.S. Thesis, California Polytechnic State University, San Luis Obispo. DOI: https://doi.org/10.15368/theses.2017.19.
8. Kennedy N.E. (2013), Seismic Design Manual for Interlocking Compressed Earth Blocks, M.S. Thesis, California Polytechnic State University, San Luis Obispo. DOI: https://doi.org/10.15368/theses.2013.85.
9. Elvin A. and Uzoegbo H. (2011), "Response of a Full-Scale Dry-Stack Masonry Structure Subject to Experimentally Applied Earthquake Loading", Journal of the South African Institution of Civil Engineering, Vol. 53, No. 1, Pp. 22–32. Available at: https://scielo.org.za (SciELO South Africa); no DOI assigned by the publisher.
10. Keshav L., Srisanthi V.G. and Balamurugan G. (2014), "Shaking Table Study of Two Reduced-Scale Single-Storey HCSE Block Masonry Building Models", Arabian Journal for Science and Engineering, Vol. 39, Pp. 2551–2560. DOI: https://doi.org/10.1007/s13369-013-0893-6.
11. Elmorsy M., Adamopoulou D., Zhang Y., Shea K., Jones L., Anastasopoulos I. and Vassiliou M.F. (2026), "Shear Behavior and Centrifuge Seismic Modeling of a Sustainable Corbel Dwelling with Interlocking Blocks", Construction and Building Materials, Vol. 515, Art. 145553. DOI: https://doi.org/10.1016/j.conbuildmat.2026.145553.
12. Silva R.A., Soares E., Oliveira D.V., Miranda T., Cristelo N.M. and Leitão D. (2015), "Mechanical Characterisation of Dry-Stack Masonry Made of CEBs Stabilised with Alkaline Activation", Construction and Building Materials, Vol. 75, Pp. 349–358. DOI: https://doi.org/10.1016/j.conbuildmat.2014.11.038.
13. Wang T., Wang Y., Zeng G., Zhang J. and Shi D. (2021), "Stability Coefficient of Interlocking Compressed Earth Block Masonry under Axial Compression", Canadian Journal of Civil Engineering, Vol. 48, No. 4, Pp. 389–398. DOI: https://doi.org/10.1139/cjce-2019-0281.
14. Sokairge H., Rashad A. and Elshafie H. (2017), "Behavior of Post-Tensioned Dry-Stack Interlocking Masonry Walls under Out of Plane Loading", Construction and Building Materials, Vol. 133, Pp. 348–357. DOI: https://doi.org/10.1016/j.conbuildmat.2016.12.071.
15. Xie G., Zhang X., Hao H., Shi T., Cui L. and Thomas J. (2023), "Behaviour of Reinforced Mortarless Interlocking Brick Wall under Cyclic Loading", Engineering Structures, Vol. 283, Art. 115890. DOI: https://doi.org/10.1016/j.engstruct.2023.115890.
16. Koudje B. and Adjovi E. (2025), "Numerical Simulation of a Shear Wall Model in Interlocking Masonry with Dry Vertical and Horizontal Joints in Compressed Earth Blocks", Buildings, Vol. 15, Art. 627. DOI: https://doi.org/10.3390/buildings15040627.
17. Zahra T. and Dhanasekar M. (2018), "Characterisation and Strategies for Mitigation of the Contact Surface Unevenness in Dry-Stack Masonry", Construction and Building Materials, Vol. 169, Pp. 612–628. DOI: https://doi.org/10.1016/j.conbuildmat.2018.03.002.

18. Hossain M.A., Totoev Y.Z. and Masia M.J. (2020), "Experimental Investigation of Frictional Behavior of Mortarless Surface in Semi-Interlocking Masonry under Cyclic Displacement", Journal of Materials in Civil Engineering, Vol. 32, No. 9, Art. 04020259. DOI: https://doi.org/10.1061/(ASCE)MT.1943-5533.0003333.

19. Buckingham E. (1914), "On Physically Similar Systems; Illustrations of the Use of Dimensional Equations", Physical Review, Vol. 4, Pp. 345–376. DOI: https://doi.org/10.1103/PhysRev.4.345.

20. Harris H.G. and Sabnis G.M. (1999), Structural Modeling and Experimental Techniques, 2nd Edn., CRC Press, Boca Raton. DOI: https://doi.org/10.1201/9780367802295. ISBN 9780849324697.

21. Zonta D., Zanardo G. and Modena C. (2001), "Experimental Evaluation of the Ductility of a Reduced-Scale Reinforced Masonry Building", Materials and Structures, Vol. 34, No. 244, Pp. 636–644. DOI: https://doi.org/10.1007/BF02482131.

22. Bourahla N. and Blakeborough A. (2015), "Similitude Distortion Compensation for a Small Scale Model of a Knee Braced Steel Frame", American Journal of Engineering and Applied Sciences, Vol. 8, No. 4, Pp. 481–488. DOI: https://doi.org/10.3844/ajeassp.2015.481.488.

23. Karaker L. and Damcı E. (2026), "Feasibility-Oriented Assessment of Physical Modeling Approaches for Shaking Table Testing of RC Minarets", Applied Sciences, Vol. 16, No. 2, Art. 990. DOI: https://doi.org/10.3390/app16020990.

24. Gumaste K.S., Nanjunda Rao K.S., Venkatarama Reddy B.V. and Jagadish K.S. (2007), "Strength and Elasticity of Brick Masonry Prisms and Wallettes under Compression", Materials and Structures, Vol. 40, No. 2, Pp. 241–253. DOI: https://doi.org/10.1617/s11527-006-9141-9.

25. Petry S. and Beyer K. (2014), "Influence of Boundary Conditions and Size Effect on the Drift Capacity of URM Walls", Engineering Structures, Vol. 65, Pp. 76–88. DOI: https://doi.org/10.1016/j.engstruct.2014.01.048.

26. Ruiz D.M., Reyes J.C., Alvarado Y.A., Vacca H., Tarque N. and Jerez S. (2025), "Investigating Scale Effects on Experimental Shear Strength of Earthen Walls (Adobe and Rammed-Earth)", Buildings, Vol. 15, Art. 689. DOI: https://doi.org/10.3390/buildings15050689.

27. Bureau of Indian Standards (2023), IS 1725:2023: Stabilized Soil Blocks Used in General Building Construction, Specification, BIS, New Delhi.

28. Bureau of Indian Standards (1993, reaffirmed), IS 13827:1993: Improving Earthquake Resistance of Earthen Buildings, Guidelines, Earthquake Engineering Sectional Committee CED 39, BIS, New Delhi.

DOI - 10.68648/ijte-q-2026-49-03-006

Comparative Study of Autoclaved Aerated Concrete (AAC) Blocks and Conventional Concrete Blocks for Sustainable Construction

Abhishek Kumar, Ravi Ranjan Kumar
Ishant Kumar
Department of Civil Engineering
Darbhanga College of Engineering
Darbhanga, Bihar
✉ abhishekkumarsinha004@gmail.com
✉ rsingh84065@gmail.com
✉ ishan.bit.2013@gmail.com

Prabhakar Kumar
Department of Mechanical Engineering
Government Engineering College
Khagaria Bihar
✉ prabhakarkumar168@yahoo.com

ABSTRACT

Rapid urbanisation demands walling materials that are simultaneously energy-efficient, low-carbon and affordable. This paper presents a multi-criteria comparison of Autoclaved Aerated Concrete (AAC) blocks with conventional solid and hollow concrete blocks across physical, mechanical, thermal, acoustic, structural, economic and environmental indicators, benchmarked against IS 2185 (Parts 1 and 3), IS 6041 and IS 6441. Compressive strength, thermal conductivity and capillary sorptivity were determined on fifteen AAC and fifteen concrete block specimens at the DCE Darbhanga laboratory, while the remaining parameters were compiled from Indian Standard limits, manufacturer data and published literature. Although concrete blocks develop higher compressive strength (5.0–15.0 N/mm² against 3.0–4.5 N/mm² for AAC), AAC masonry reduces wall self-weight by about 70 per cent, jointing mortar consumption by 75 per cent and thermal transmittance by 76 per cent (0.78 against 3.25 W/m²K). For a ten-storey commercial building modelled in Seismic Zone III, AAC infill lowered structural steel by 8.5 per cent and concrete volume by 6.8 per cent, giving a 10.2 per cent reduction in total structural cost, while EnergyPlus simulation indicated an 18.3 per cent annual cooling energy saving. Embodied carbon is 54 per cent lower per square metre of wall. An engineering decision matrix is proposed to guide material selection.

KEYWORDS *: Autoclaved aerated concrete; Concrete masonry units; Thermal transmittance; Embodied carbon; Life-cycle cost.*

INTRODUCTION

Growing Need for Sustainable Construction Materials

The global construction industry accounts for approximately 36 per cent of the world's final energy consumption, 39 per cent of energy-related carbon dioxide (CO_2) emissions and nearly 40 per cent of virgin raw material extraction. In developing countries such as India, rapid population growth and urban migration drive the construction of millions of square metres of residential, commercial and industrial floor space every year. Masonry walling constitutes 30 to 45 per cent of the total dead load in multi-storey RCC framed structures, the proportion varying with building height, column spacing and wall-to-floor area ratio [1]. It also governs both the thermal performance of the building envelope and the quality of the indoor environment.

Traditionally, fired clay bricks and conventional concrete masonry units (CMUs), both solid and hollow, have dominated masonry envelope systems. However, the production of fired bricks consumes fertile topsoil and burns high-sulfur coal in uncontrolled kilns, causing significant atmospheric pollution and soil

degradation [2]. Although conventional concrete blocks reduce the loss of agricultural land, their production depends heavily on ordinary Portland cement (OPC), the calcination of which emits approximately 0.85 to 0.90 tonnes of CO_2 per tonne of clinker. In addition, the high density of these units increases earthquake-induced base shear and therefore demands heavier RCC foundations, columns and beams [3]. Consequently, cellular and lightweight masonry technologies that decouple environmental degradation from structural performance are attracting increasing attention from civil and structural engineers.

Existing research, however, concentrates largely on isolated parameters — thermal, mechanical or economic — and lacks a thorough, India-specific synthesis across physical, structural, economic and environmental dimensions under Indian Standard specifications [4], [5]. By offering a unified, multi-criteria comparative framework to guide material selection for framed and load-bearing masonry applications, the present study addresses that gap.

Background on Conventional Concrete Blocks

According to IS 2185 (Part 1): 2005 [6], conventional concrete masonry blocks are produced from a consolidated mixture of cement, coarse aggregate (6 to 10 mm graded gravel), fine aggregate (river sand or manufactured M-sand) and water, compacted using mechanical vibratory presses and water-cured for 14 to 28 days [7]. Three principal typologies are recognised:

1. Solid concrete blocks: dry densities of 1800 to 2200 kg/m³, compressive strengths of 4.0 to 15.0 N/mm², and a solid volume exceeding 75 per cent of the gross block volume. They are traditionally used in load-bearing external walls, retaining walls and heavy foundation masonry.
2. Hollow concrete blocks (Grades A and B, load-bearing): internal voids of 25 to 50 per cent with a minimum web thickness of 25 mm, densities of 1500 to 1800 kg/m³ and compressive strengths of 3.5 to 7.0 N/mm², offering marginally improved thermal insulation relative to solid blocks owing to enclosed air pockets.
3. Grade C non-load-bearing hollow blocks: intended strictly for internal partitions, with densities of 1000 to 1500 kg/m³ and compressive strengths exceeding 1.5 N/mm².

Despite their proven structural capacity, conventional concrete blocks exhibit significant thermal bridging ($k \approx 1.10–1.75$ W/m·K), require thick mortar bedding joints (10 to 15 mm), impose substantial structural dead load (2.8 to 4.2 kN/m² for 200 mm masonry) and entail considerable on-site labour and water curing [8].

AAC Blocks: Composition and Autoclaving Process

Autoclaved Aerated Concrete, originally patented in 1924 by the Swedish architect Axel Eriksson, is an ultra-lightweight, cellular, precast building material governed in India by IS 2185 (Part 3): 1984 (Reaffirmed 2020) [9]. It is synthesised from a finely pulverised slurry of the following constituents:

- Siliceous material: Class F pulverised fuel ash conforming to IS 3812 [10], comprising 65 to 70 per cent of the total dry solid mass, or quartz sand.
- Calcareous binders: ordinary Portland cement (OPC 53 grade, 15–20 per cent) and quicklime (CaO, 10–15 per cent).
- Sulfate setting regulator: gypsum ($CaSO_4 \cdot 2H_2O$) or anhydrite ($CaSO_4$), 2–4 per cent.
- Blowing agent: finely divided aluminium powder (0.05–0.08 per cent by weight of solids).
- Water: potable water at a controlled water-to-solid ratio of 0.55 to 0.65.

During slurry mixing and mould casting, an exothermic reaction occurs between the finely ground metallic aluminium powder and the calcium hydroxide liberated by cement and lime hydration:

$$2Al + 3Ca(OH)_2 + 6H_2O \rightarrow 3CaO \cdot Al_2O_3 \cdot 6H_2O + 3H_2 \text{ (gas)} \quad (1)$$

The evolution of microscopic hydrogen bubbles expands the cake volume by approximately 200 to 300 per cent, creating an evenly dispersed, closed-cell, spherical micro-porosity with pore sizes ranging from 50 μm to 1.5 mm. Following pre-curing and wire cutting to high dimensional tolerances (±1.5 mm), the green blocks are transferred to an autoclave chamber, where they undergo high-pressure hydrothermal curing at 1.2 to 1.4 MPa and 180 to 200 °C for 10 to 12 hours. This hydrothermal

mineralogy imparts exceptional dimensional stability, low drying shrinkage (below 0.04 per cent), superior fire resistance and a remarkable strength-to-density ratio [11]. The complete manufacturing sequence is shown in Fig. 1.

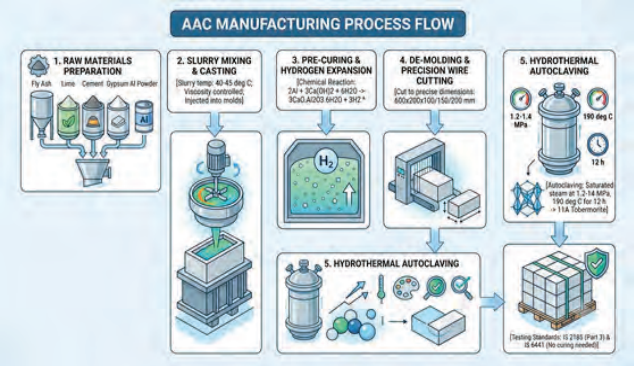


Fig. 1. AAC manufacturing process flow

Objectives and Scope of the Study

The principal objectives of this study are:

1. To determine whether AAC blocks provide advantages in strength-to-weight ratio, thermal insulation and life-cycle cost that offset their lower compressive strength for non-load-bearing infill applications in framed structures, and to identify the conditions under which each material system is optimal.
2. To compile and verify baseline performance parameters from laboratory testing, published experimental studies, manufacturer data and Indian Standard specifications.
3. To establish whether the higher unit cost of AAC blocks is offset by reductions in mortar, plastering, structural steel and HVAC operational costs over the building life cycle, and to quantify the net economic advantage for a representative 100 m^2 wall assembly.
4. To evaluate life-cycle environmental indicators, including embodied carbon (kg CO_2e/m^2), industrial waste diversion and green building certification credits under GRIHA v2019 and LEED v4.1.
5. To synthesise an engineering application decision matrix delineating structural and non-structural deployment domains.

LITERATURE REVIEW

Research comparing autoclaved cellular concrete with conventional cementitious masonry units spans four decades and may be grouped under four themes: mix development and the valorisation of secondary materials, thermal and durability performance, comparative sustainability assessment, and the structural behaviour of masonry assemblages.

Mix Development and Valorisation of Secondary Materials

A substantial body of work has examined the replacement of virgin constituents in cellular and conventional concrete units with industrial residues. Wang et al. [12] prepared non-autoclaved aerated concrete from the pyrolysis residues of oil-based drilling cuttings, confirming that silica-rich industrial wastes can substitute for primary siliceous feedstock without compromising the cellular structure. Poon et al. [13] demonstrated that recycled aggregates can be incorporated into moulded concrete bricks and blocks at moderate replacement levels with acceptable strength loss, while Tam et al. [7], reviewing recycled aggregate concrete over the period 2000 to 2017, found replacement ratios of up to 30 per cent generally tolerable for non-structural applications. These studies establish the technical basis for the high fly ash content that distinguishes AAC from conventional concrete blocks, but they address the constituent material rather than the performance of the completed wall.

Thermal and Durability Performance

Mathey and Rossiter [14], in an early and still widely cited review, concluded that AAC combines low thermal conductivity with adequate load-bearing capacity, rendering it suitable for both structural and non-structural use, although supplementary insulation may be required in severe climates. Chaipanich and Chindaprasirt [15] examined the properties and durability of AAC masonry blocks in detail, reporting that the closed-cell pore structure governs moisture transport and that prolonged saturation measurably reduces compressive strength. Namsone et al. [11] observed comparable durability behaviour in high-performance foamed concrete, confirming that the cellular pore system rather than the binder chemistry

controls freeze-thaw and moisture response. Kubilay et al. [16] provided high-resolution field measurements of wind-driven rain on low-rise buildings, quantifying the driving-rain exposure against which the capillary behaviour of porous walling units must be assessed.

Comparative Sustainability Assessment

Gautam and Saxena [17] compared AAC blocks with fired red bricks and reported significant reductions in embodied energy and dead load, although the analysis was confined to material-level indicators. Kamal [8] extended the assessment to the sustainability potential of AAC in building envelopes, and Dey et al. [4], in a recent state-of-the-art review, consolidated four decades of AAC research and identified life-cycle assessment and region-specific cost modelling as the principal outstanding gaps. Azhar et al. [5] addressed part of that gap through a climate-specific meta-analysis of global AAC performance, showing that decarbonisation outcomes vary substantially with climate zone and grid carbon intensity. Oladiran and Simeon [18] and Rogers Simeon and Oladiran [19] surveyed awareness, barriers and improvement strategies for AAC adoption in Nigeria and South Africa, concluding that limited comparative performance data, rather than technical inadequacy, constrains uptake in developing economies.

Structural Behaviour of Masonry Assemblages

Cheng et al. [2] compared the in-plane shear performance of masonry walls strengthened with different fibre-reinforced materials, establishing reference shear capacities for unreinforced and strengthened panels. Sreekeshava et al. [20] investigated the strengthening of masonry wallettes using a polyester geo-fabric reinforced cementitious matrix, and Chitra and Mohan [3] examined the ductility behaviour of reinforced concrete beam-column joints, which governs the seismic demand transmitted to and by infill masonry. Collectively, these studies indicate that infill mass and stiffness, rather than unit compressive strength alone, dominate the seismic response of framed structures.

Research Gap

The literature consistently confirms the thermal, environmental and dead-load advantages of AAC, yet the evidence remains fragmented. Material-level studies do not extend to wall assemblies; thermal studies rarely quantify the structural or foundation consequences of reduced mass; and sustainability assessments seldom adopt Indian Standard specifications or Indian cost structures. No published study simultaneously quantifies mechanical, thermal, structural, economic and cradle-to-gate environmental outcomes for AAC and conventional concrete blocks under IS 2185, IS 6041 and IS 6441. The present work addresses that gap through an integrated, multi-criteria comparison supported by laboratory testing, structural modelling and life-cycle costing.

MATERIALS AND METHODS

Raw Material Composition and Manufacturing Parameters

Compressive strength, thermal conductivity and capillary sorptivity were determined by original laboratory testing on fifteen AAC block samples and fifteen solid concrete block samples at DCE Darbhanga laboratory operating in accordance with ISO/IEC 17025 [21]. Data for water absorption, drying shrinkage, fire resistance and acoustic transmission were compiled from IS 2185 specification limits, manufacturer technical datasheets and peer-reviewed studies, because these parameters are standardised and exhibit little inter-manufacturer variability.

The AAC samples were manufactured in June 2026 and randomly selected from a production batch of 5,000 blocks to ensure representativeness. The sample size was fixed at fifteen against the minimum of ten specimens required by IS 6441 (Part 5) [22], so that variability could also be analysed statistically. Specimens were cured for 28 days and conditioned to a moisture content of 10 ± 2 per cent as required by IS 2185 (Part 3) before testing. Solid concrete block samples were manufactured in June 2026 and water-cured for 28 days in accordance with IS 2185 (Part 1), and ten hollow concrete block samples were obtained from a comparable source. All samples conformed to to the nominal dimensions given in Table 1, with dimensional tolerances verified as per IS 2185 [6]. The comparative composition and manufacturing parameters of the three unit types are summarised in Table 1.

Table 1. Raw material composition and manufacturing parameters of the masonry units

Parameter / Constituent	AAC Block	Solid Concrete Block (SCB)	Hollow Concrete Block (HCB, Gr A)
Primary binder	OPC 53 grade (15–20%) + lime (10–15%)	OPC 43/53 grade (10–14%)	OPC 43/53 grade (10–14%)
Pozzolan / aggregate	Class F fly ash (65–70%) as per IS 3812	Graded river/M-sand + 6–10 mm crushed stone	Graded sand + 6–10 mm crushed gravel
Aerating agent	Aluminium powder (0.05–0.08%)	None (compacted dense concrete)	None (moulded with internal cores)
Sulfate source	Gypsum / anhydrite (2–4%)	None / gypsum in cement	None / gypsum in cement
Manufacturing method	Chemical expansion + wire cutting + autoclaving	Heavy vibro-compaction, hydraulic press	Heavy vibro-compaction with core mould
Curing regime	Hydrothermal steam (1.2–1.4 MPa, 190 °C, 12 h)	Atmospheric water ponding/ spray (14–28 days)	Water spray / steam curing (14–28 days)
Standard dimensions	600 × 200 × (100/150/200) mm	400 × 200 × (150/200) mm	400 × 200 × (150/200) mm
Dimensional tolerance	±1.5 mm length/height, ±1.0 mm width	±3.0 to ±5.0 mm	±3.0 to ±5.0 mm

Testing Protocols and Governing Standards

The selected protocols address the multi-criteria research question: compressive strength verifies IS 2185 compliance and enables calculation of the strength-to-density ratio; thermal conductivity quantifies envelope energy performance; water absorption and capillary sorptivity assess durability under driving rain; and drying shrinkage evaluates crack risk in thin-bed jointed masonry [15]. Taken together, these parameters determine suitability for framed infill applications in tropical and composite climates. Testing followed IS 6441 (Parts 1 to 9): 1972 and IS 2185 (Part 1): 2005, with the international AAC specification ASTM C1693 [23] consulted for cross-reference.

- Oven-dry density (IS 6441, Part 1): specimens (150 × 150 × 150 mm prisms for AAC; full-size units for concrete blocks) were conditioned in a ventilated oven at 105 ± 5 °C until constant mass was achieved.
- Compressive strength (IS 6441, Part 5 and IS 2185, Part 1): measured on a compression testing machine calibrated to 2000 kN. Specimens were loaded perpendicular to the direction of rise at a constant stress rate of 0.05 to 0.10 N/mm²/s until ultimate crushing. Moisture content of AAC specimens was controlled at 10 ± 2 per cent by weight at the time of testing, as required by IS 2185 (Part 3).
- Water absorption and moisture movement: a 24-hour cold water immersion test was carried out to measure total volumetric absorption, and capillary sorptivity was assessed concurrently in accordance with ASTM C1585 [24].
- Thermal conductivity (k-value): determined using a guarded hot plate apparatus conforming to ASTM C177 [25] and IS 3792 [26] under steady-state heat flux at a mean specimen temperature of 25 °C.
- Drying shrinkage (IS 6441, Part 2): measured by recording the initial reference length of fully saturated specimens with an extensometer dial gauge and tracking longitudinal contraction to a constant dry state.

COMPARATIVE ANALYSIS

Table 2 presents the verified experimental data alongside the standardised Indian Standard specification limits for the three masonry types.

Table 2. Comparative performance of AAC, solid concrete and hollow concrete blocks

Property / Parameter	AAC Blocks	Solid Concrete Blocks (SCB)	Hollow Concrete Blocks (HCB, Gr A)
Governing Indian code	IS 2185 (Part 3): 1984	IS 2185 (Part 1): 2005	IS 2185 (Part 1): 2005
Oven-dry density	550–650 kg/m³	1800–2200 kg/m³	1500–1800 kg/m³
Compressive strength	3.0–4.5 N/mm²	5.0–15.0 N/ mm²	3.5–7.0 N/mm²
Thermal conductivity, k	0.16–0.20 W/m·K	1.10–1.75 W/m·K	0.85–1.15 W/m·K

U-value (200 mm wall)	0.78 W/m²·K	3.25 W/m²·K	2.62 W/m²·K
Sound insulation (STC)	40–44 dB	48–50 dB	44–46 dB
Water absorption (24 h)	25–35% by dry mass	6–10% by dry mass	8–12% by dry mass
Capillary sorptivity	0.82 kg/ m²·min$^{0.5}$	1.45 kg/ m²·min$^{0.5}$	1.60 kg/ m²·min$^{0.5}$
Unit weight per block	≈13.8 kg (600 × 200 × 200 mm)	≈32.8 kg (400 × 200 × 200 mm)	≈22.7 kg (400 × 200 × 200 mm)
Wall self-weight (200 mm)	1.15–1.25 kN/ m²	3.80–4.20 kN/ m²	2.80–3.20 kN/ m²
Fire resistance rating	4.0 h (150–200 mm)	2.0–3.0 h	1.5–2.0 h
Dimensional tolerance	±1.5 mm	±3.0 to ±5.0 mm	±3.0 to ±5.0 mm
Mortar type / bed thickness	Thin-bed polymer adhesive (2–3 mm)	Cement–sand 1:6 (10–15 mm)	Cement–sand 1:6 (10–15 mm)
Laying speed (mason team)	18–22 m²/day	8–10 m²/day	10–12 m²/day
Plastering requirement	3–4 mm skim coat / gypsum plaster	12–15 mm internal + 18 mm external	12–15 mm internal + 18 mm external
On-site water curing	Nil	7–14 days	7–14 days
Fly ash utilisation	65–70% by dry mass	0–15%	0–15%
Embodied carbon (200 mm wall)	29–35 kg CO_2e/m²	64–76 kg CO_2e/m²	48–58 kg CO_2e/m²

Detailed Analysis of Key Parameters

Density and Self-Weight Reduction

The dry density of AAC blocks (550 to 650 kg/m³) is approximately one-third to one-fourth that of solid concrete blocks (1800 to 2200 kg/m³) and less than half that of hollow concrete blocks (1500 to 1800 kg/m³). For a 200 mm wall, AAC reduces masonry self-weight from about 4.0 kN/m² to about 1.2 kN/m², a reduction of roughly 70 per cent. In the ten-storey building analysed under Structural Superstructure Savings, this dead load reduction produced a 22.4 per cent reduction in seismic base shear when modelled in ETABS as per IS 1893: 2016 [27], which in turn permitted reductions in column and foundation dimensions. The magnitude of these effects is site- and design-specific and will vary with building height, column spacing and soil conditions.

Compressive Strength and Infill Performance

While solid concrete blocks provide the higher raw compressive strengths (5.0 to 15.0 N/mm²) required for unreinforced load-bearing walls up to G+3, framed high-rise towers require only non-load-bearing infill. As per IS 6041 [28], IS 1905 [29] and NBC 2016, a compressive strength of 3.0 N/mm² is adequate for infill masonry. AAC blocks, at 3.0 to 4.5 N/mm², comfortably meet this threshold while satisfying wind and impact criteria.

Thermal Insulation and HVAC Savings

The thermal conductivity of AAC (k ≈ 0.18 W/m·K) is roughly one-eighth that of solid concrete (k ≈ 1.45 W/m·K). The resulting wall thermal transmittance is 0.78 W/m²·K for AAC against 3.25 W/m²·K for solid concrete. To quantify the operational benefit, a ten-storey office building (12,000 m² built-up area, 40 per cent window-to-wall ratio, composite climate as per ECBC [30], HVAC COP of 3.2, occupancy 08:00 to 18:00) was modelled in EnergyPlus 9.5. Replacing 200 mm solid concrete walls with AAC walls reduced annual cooling energy by 18.3 per cent, from 142 kWh/ m² to 116 kWh/m². Varying the window-to-wall ratio and roof insulation level across the range encountered in Indian commercial practice gives savings of 12 per cent (30 per cent WWR with high roof insulation) to 24 per cent (50 per cent WWR with low roof insulation). In many applications, AAC walls satisfy ECBC envelope requirements without additional external insulation.

Water Absorption and Capillary Sorptivity

Total immersion absorption is high (25 to 35 per cent) because of open-surface macropores, but it is capillary sorptivity that governs performance under ambient driving rain. The closed spherical internal pores of AAC give a very low capillary sorptivity (0.82 kg/m²·min$^{0.5}$ against 1.45 to 1.60 kg/m²·min$^{0.5}$ for concrete blocks), which, when detailed with breathable external coatings, effectively protects against dampness and efflorescence under the driving-rain exposures measured for low-rise buildings by Kubilay et al. [16].

Fire Resistance

AAC carries a certified four-hour fire endurance rating at 150 to 200 mm thickness and is non-combustible (Euroclass A1) under IS 2185 (Part 3). Dense concrete blocks achieve two to three hours but are prone to explosive surface spalling under abrupt thermal shock.

Dimensional Precision and Mortar Usage

Wire-cut AAC blocks with a tolerance of ±1.5 mm permit thin-bed polymer adhesive joints. This eliminates the requirement for ten days of on-site water curing, reduces mortar consumption by 75 to 80 per cent, and allows direct application of a 3 to 4 mm skim coat or gypsum plaster in place of thick sand–cement plaster.

COST ANALYSIS

An itemised cost analysis was carried out for 100 m² of completed masonry walling of 200 mm nominal thickness at prevailing commercial construction rates in the Indian market, converted at ₹83 per US dollar. The results are presented in Table 3.

Table 3. Itemised direct cost comparison for 100 m² of finished 200 mm masonry wall

Cost Component	AAC Blocks (thin-bed adhesive)	Solid Concrete Blocks (1:6 mortar)
1. Block units (base material)	₹68,000 (US$819) [860 blocks @ ₹3,300/m³]	₹55,000 (US$663) [1,312 blocks @ ₹42/block]
2. Jointing mortar / adhesive	₹9,600 (US$116) [36 bags polymer adhesive @ ₹265]	₹18,500 (US$223) [28 bags cement + 5.2 t sand]
3. Masonry labour and laying	₹18,000 (US$217) [5 days @ 20 m²/day]	₹32,000 (US$386) [11 days @ 9 m²/day]
4. Wall plastering (int. + ext.)	₹28,000 (US$337) [thin skim coat + rendering]	₹58,000 (US$699) [15 mm internal + 18 mm external]
5. Water curing and cleaning	₹1,500 (US$18) [no curing required]	₹8,000 (US$96) [14 days continuous curing]
Total direct walling cost	₹1,25,100 (US$1,507)	₹1,71,500 (US$2,066)
Direct saving with AAC	₹46,400 per 100 m² (27.06%)	—

Structural Superstructure Savings

A typical ten-storey commercial office building was modelled in ETABS v19 with the following parameters: plan dimensions 40 m × 30 m (1,200 m² per floor), floor-to-floor height 3.5 m, regular column grid of 6 m × 6 m (7 × 6 bays), Fe500 reinforcement, slab thickness 150 mm, live load 3 kN/m² for office occupancy as per IS 875 (Part 2) [31], floor finish 1.5 kN/m², Seismic Zone III, medium soil (Type II, Z = 0.16) as per IS 1893: 2016, importance factor 1.0, response reduction factor R = 5, and response spectrum analysis as per IS 1893 Clause 6.4.

The baseline configuration was selected to represent typical urban commercial construction in composite-climate cities such as Delhi, Jaipur and Pune, in line with NBC 2016 guidance, a 6 m grid being standard for office buildings with a 3.5 m floor-to-floor height. Zone III denotes moderate seismicity and covers a large share of India's urban population. Two infill scenarios were modelled: (i) 200 mm solid concrete block masonry (unit weight 20 kN/m³, wall load 4.0 kN/m per floor height) and (ii) 200 mm AAC block masonry (unit weight 6.5 kN/m³, wall load 1.3 kN/m per floor height). Load combinations were taken from Table 18 of IS 456: 2000 [32]. Preliminary column sizes ranged from 450 × 450 mm at ground floor to 300 × 300 mm at the top floor, with beams of 300 × 450 mm. The principal outcomes were as follows.

- Dead load: the AAC scheme reduces total building dead weight by 3,200 tonnes.
- Reinforcement steel: column and beam bending moments and axial forces fall by 12 to 16 per cent, giving an 8.5 per cent reduction in total structural steel across the frame (approximately 38 tonnes saved at ₹65,000 per tonne, a saving of ₹24,70,000).
- Concrete volume: reduced column and footing cross-sections give a 6.8 per cent reduction in M30/M40 concrete volume (approximately 210 m³ saved at ₹5,200 per m³, a saving of ₹10,92,000).
- Foundations: for weak soils, pile group quantities and depths reduce by 14 to 18 per cent, worth approximately ₹18,00,000 for this building.

Weighting the material savings by their cost shares —

structural steel at 55 per cent of superstructure cost with 8.5 per cent savings, RCC concrete at 35 per cent with 6.8 per cent savings, and formwork and labour at 10 per cent unchanged — gives a weighted superstructure saving of 7.1 per cent. Including the foundation savings, the total structural cost reduction is 10.2 per cent. Repeating the analysis across Seismic Zones II to V and across soft, medium and hard soil conditions gives structural savings in the range of 8 to 16 per cent, confirming that outcomes are site-specific and depend on seismic demand and foundation conditions. This spread is consistent with the climate- and context-dependence of AAC performance reported by Azhar et al. [5].

Operational Life-Cycle Costing and HVAC Energy Amortisation

Owing to an envelope U-value of 0.78 W/m²·K for AAC against 3.25 W/m²·K for solid concrete blocks, the peak chiller cooling load for typical conditioned office space falls by approximately 18 to 22 tonnes of refrigeration per 1,000 m² of external wall, saving about ₹9,00,000 per 1,000 m² of façade in initial HVAC sizing. Expressed per unit of wall area, the annual cooling energy reduction averages 16.5 kWh per m² of external wall, equivalent to the 26 kWh per m² of floor area reported under Thermal Insulation and HVAC Savings for the same building.

A 50-year operational horizon was evaluated assuming a real discount rate of 6 per cent per annum, consistent with Indian public infrastructure appraisal practice; energy price escalation of 3 per cent per annum, based on 2010–2020 average commercial electricity tariff trends in composite-climate cities; maintenance at 0.5 per cent of initial wall capital per annum; and HVAC equipment replacement in years 15 and 30 at 60 per cent of initial HVAC capital cost. Under these assumptions, the present-value energy saving of ₹1,845 per m² of AAC wall over 50 years exceeds the initial material cost premium of ₹540 per m² by a factor of 3.4. Actual savings will vary with energy tariffs, climate and building operation practices.

SUSTAINABILITY AND ENVIRONMENTAL IMPACT

Industrial Byproduct Utilisation

Thermal coal-fired power stations generate millions of tonnes of pulverised fuel ash annually, which occupies valuable agricultural land in containment slurry ponds and poses severe groundwater leaching hazards [17]. AAC blocks contain 65 to 70 per cent fly ash by dry solid weight. A single production facility generating 300,000 m³ of AAC blocks per year consumes approximately 135,000 tonnes of fly ash, directly repurposing industrial waste into durable civil infrastructure. Conventional concrete blocks, by contrast, typically contain 0 to 15 per cent fly ash and rely primarily on virgin-mined aggregates.

Embodied Energy and Carbon Footprint

Table 4. Cradle-to-gate life-cycle inventory of the three masonry unit types

Life-Cycle Metric	AAC	Solid Concrete Block (SCB)	Hollow Concrete Block (HCB)
Embodied energy (MJ/m³)	1,050–1,250	1,800–2,200	1,400–1,650
Embodied energy (MJ/m² of 200 mm wall)	210–250	360–440	280–330
Embodied carbon (kg CO_2e/m^3)	145–175	320–380	240–290
Embodied carbon (kg CO_2e/m^2 of wall)	29–35	64–76	48–58
Virgin aggregate consumption (kg/m²)	0	340	220
Potable water for curing (L/m²)	0 (closed factory loop)	65–80	50–65

A cradle-to-gate life cycle assessment was carried out as per ISO 14044 [33], with the system boundary extending from raw material extraction to the factory gate, a functional unit of 1 m² of finished 200 mm wall including jointing mortar or adhesive and 12 mm plaster, and GWP_{100} characterisation factors taken from Ecoinvent 3.8 [34]. On this basis, AAC masonry (32 kg CO_2e/m^2) exhibits 54 per cent lower embodied carbon than solid concrete masonry (70 kg CO_2e/m^2)

under baseline assumptions of the Indian electricity grid mix, a 100 km transport distance and OPC 53 cement. Sensitivity analysis shows the reduction to be robust, ranging from 48 to 62 per cent depending on grid carbon intensity (0.82 to 0.95 kg CO_2e/kWh) and transport distance (50 to 200 km).

The reduction is attributable to three factors: the substitution of uncalcined fly ash (65 to 70 per cent by weight) for virgin cement and aggregate; a 65 per cent decrease in mass per m² of wall; and a 75 per cent decrease in jointing mortar consumption under thin-bed adhesive. The comparison assumes equivalent wall area and does not account for variations in transport distance, on-site construction energy or end-of-life recycling potential, all of which may differ with project location.

Green Building Certification Credits

- GRIHA v2019: contributes directly to Criterion 6 (soil preservation), Criterion 9 (sustainable building materials, above 60 per cent fly ash), Criterion 13 (ECBC envelope optimisation) and Criterion 21 (zero on-site curing water).
- LEED v4.1: eligible for the EQ credit (indoor thermal and acoustic comfort), the EA credit (optimise energy performance under ASHRAE 90.1) and the MR credit (recycled content from pre-consumer industrial waste).

RESULTS AND DISCUSSION

Synthesising the comparative parameters of Tables 2 to 4 yields three principal findings.

1. Structural efficiency. Taking mid-range values, AAC exhibits a compressive strength-to-density ratio of approximately 6.3 × 10^{-3} N·m³/(mm²·kg) (3.75 N/mm² at 600 kg/m³) against 5.0 × 10^{-3} for solid concrete blocks (10.0 N/mm² at 2000 kg/m³). AAC is therefore substantially more efficient for non-load-bearing framed infill masonry, where structural economy is governed by dead load minimisation rather than by unit strength.
2. Thermal and acoustic synergy. Whereas solid concrete blocks rely purely on inertial thermal mass and the acoustic mass law, AAC blocks provide an integrated thermal insulating envelope without additional synthetic polymer insulation layers, while retaining an STC rating adequate for internal partitions.
3. Moisture mechanics. Although AAC exhibits a high saturation capacity under hydraulic immersion testing, its low capillary sorptivity under ambient driving rain demonstrates that moisture penetration concerns are largely alleviated when correct construction practice is observed, namely damp-proof course installation, starter courses and breathable coatings.

These findings are consolidated in the engineering application decision matrix of Fig. 2, which maps each material system to the building typologies and exposure conditions for which it is best suited.

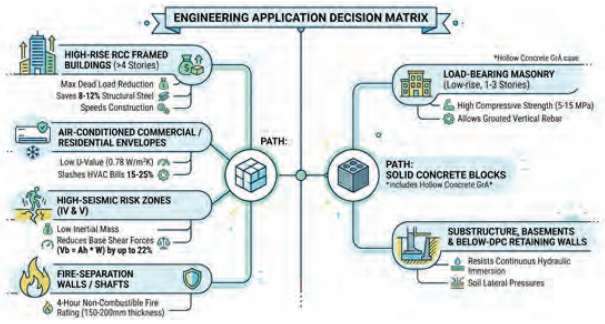


Fig. 2. Engineering application decision matrix for AAC and conventional concrete block masonry

CONCLUSIONS AND RECOMMENDATIONS

1. Physical and mechanical performance. AAC blocks have an oven-dry density of 550 to 650 kg/m³, approximately 70 per cent lighter than solid concrete blocks, reducing 200 mm wall self-weight from about 4.0 kN/m² to about 1.2 kN/m². Although conventional concrete blocks provide higher raw compressive strength (5.0 to 15.0 N/mm² against 3.0 to 4.5 N/mm² for AAC), AAC meets and exceeds the minimum requirements of IS 2185 (Part 3) for framed infill.
2. Thermal and energy performance. A thermal conductivity of 0.18 W/m·K lowers wall thermal transmittance to 0.78 W/m²·K, a 76 per cent reduction relative to uninsulated concrete blockwork, cutting annual cooling energy demand by 18.3 per cent in the modelled building and by

12 to 24 per cent across the range of envelope configurations examined.

3. Economic life-cycle viability. Despite a 15 to 25 per cent higher unit purchase price, AAC masonry achieves an approximately 27 per cent direct cost saving per 100 m² of finished wall (Table 3) under the assumed cost structure, driven by a 75 per cent reduction in jointing mortar consumption, elimination of thick sand–cement plastering and faster installation. In multi-storey framed buildings, the accompanying dead load reduction yields an 8.5 per cent saving in structural reinforcement steel and a 6.8 per cent saving in concrete volume, giving a 10.2 per cent reduction in total structural cost once foundation savings are included. These savings are sensitive to regional variation in labour and material costs, and site-specific analysis remains necessary.

4. Environmental decarbonisation. AAC blocks reduce embodied carbon by 54 per cent per square metre of wall assembly (32 kg CO_2e/m^2 against 70 kg CO_2e/m^2), utilise 65 to 70 per cent industrial fly ash by weight, and divert thermal power waste from landfill.

Engineering recommendations

- For high-rise commercial and residential framed towers above G+3, AAC blocks with 2 to 3 mm thin-bed polymer adhesive should be adopted as the standard infill material in order to optimise structural dead load reduction, seismic shear mitigation and operational energy efficiency.
- For low-rise load-bearing structures below G+2 without RCC frames, high-strength solid or grouted hollow concrete blocks conforming to IS 2185 (Part 1) remain preferable, as they carry direct axial gravity and shear loads.
- For substructure, retaining and wet utility masonry, the high porosity and capillary structure of AAC may cause freeze–thaw damage in cold climates where moisture is trapped, efflorescence and surface degradation where rising damp is uncontrolled, and a measurable loss of compressive strength under prolonged saturation [15]. AAC may nevertheless be used in above-ground wet areas such as bathrooms and kitchens provided that a continuous damp-proof course is installed at plinth level, wet surfaces are treated with cementitious waterproofing or waterproof tile adhesive, and breathable silicate-based or acrylic external coatings are applied to allow moisture vapour to escape. Without these precautions, mould growth, plaster delamination and reduced service life may follow. Detailed construction guidance is given in Clause 6 of IS 6041: 1985.

ACKNOWLEDGEMENTS

The authors thank Darbhanga College of Engineering, Darbhanga, for laboratory facilities and technical support.

REFERENCES

1. Bureau of Indian Standards (2016), National Building Code of India 2016, Volumes 1 and 2, BIS, New Delhi.
2. S. Cheng, S. Yin and L. Jing (2020), Comparative Experimental Analysis on the In-Plane Shear Performance of Brick Masonry Walls Strengthened with Different Fibre Reinforced Materials, Construction and Building Materials, Vol. 259, 120387.
3. R. Chitra and S. J. Mohan (2021), Reinforced Concrete Beam-Column Joint's Ductility Behaviour, Materials Today: Proceedings, Vol. 51, pp. 1069-1073.
4. P. Dey, A. Paul, M. Dhar, D. Das and R. Saha (2025), State-of-the-Art Review and Future Prospects of Autoclaved Aerated Concrete for Building a Sustainable Tomorrow, Environment, Development and Sustainability, pp. 1-42.
5. J. Azhar, S. Aydin and M. S. Zami (2025), Decarbonizing Buildings with Autoclaved Aerated Concrete (AAC): Climate-Specific Meta-Analysis of Global Performance, Journal of Asian Architecture and Building Engineering.
6. Bureau of Indian Standards (2005, Reaffirmed 2021), IS 2185 (Part 1): Concrete Masonry Units - Specification - Part 1: Hollow and Solid Concrete Blocks, BIS, New Delhi.
7. V. W. Y. Tam, M. Soomro and A. C. J. Evangelista (2018), A Review of Recycled Aggregate in Concrete Applications (2000-2017), Construction and Building Materials, Vol. 172, pp. 272-292.
8. M. A. Kamal (2020), Analysis of Autoclaved Aerated Concrete (AAC) Blocks with Reference to Its Potential and Sustainability, Journal of Building Materials and Structures, Vol. 7, No. 1, pp. 76-86.

9. Bureau of Indian Standards (1984, Reaffirmed 2020), IS 2185 (Part 3): Specification for Concrete Masonry Units - Part 3: Autoclaved Cellular (Aerated) Concrete Blocks, BIS, New Delhi.

10. Bureau of Indian Standards (2013), IS 3812 (Part 1): Pulverized Fuel Ash - Specification - Part 1: For Use as Pozzolana in Cement, Cement Mortar and Concrete, BIS, New Delhi.

11. E. Namsone, G. Sahmenko and A. Korjakins (2017), Durability Properties of High Performance Foamed Concrete, Procedia Engineering, Vol. 172, pp. 760-767.

12. C. Wang, X. Lin, D. Wang, M. He and S. Zhang (2018), Utilization of Oil-Based Drilling Cuttings Pyrolysis Residues of Shale Gas for the Preparation of Non-Autoclaved Aerated Concrete, Construction and Building Materials, Vol. 162, pp. 359-368.

13. C. S. Poon, S. C. Kou and L. Lam (2002), Use of Recycled Aggregates in Moulded Concrete Bricks and Blocks, Construction and Building Materials, Vol. 16, No. 5, pp. 281-289.

14. R. G. Mathey and W. J. Rossiter (1990), A Review of Autoclaved Aerated Concrete Products, NISTIR 4331, National Institute of Standards and Technology, Gaithersburg, MD.

15. A. Chaipanich and P. Chindaprasirt (2015), The Properties and Durability of Autoclaved Aerated Concrete Masonry Blocks, in Eco-Efficient Masonry Bricks and Blocks: Design, Properties and Durability, Woodhead Publishing, Cambridge, pp. 215-230.

16. A. Kubilay, D. Derome, B. Blocken and J. Carmeliet (2014), High-Resolution Field Measurements of Wind-Driven Rain on an Array of Low-Rise Cubic Buildings, Building and Environment, Vol. 78, pp. 1-13.

17. P. Gautam and N. Saxena (2013), Comparison of Autoclaved Aerated Concrete Blocks with Red Bricks, International Journal of Engineering Research and Technology, Vol. 2, No. 10.

18. O. Oladiran and D. R. Simeon (2023), Consciousness and Prospects of Autoclaved Aerated Concrete (AAC) Blocks for Wall Construction: Comparative Study of Nigeria and South Africa, Journal of Construction Business and Management, Vol. 6, No. 2, pp. 1-10.

19. D. Rogers Simeon and O. Joseph Oladiran (2023), Barriers and Improvement Strategies for the Use of Autoclaved Aerated Concrete Blocks, Journal of Contemporary Research in the Built Environment, Vol. 7, No. 1.

20. K. S. Sreekeshava, A. S. Arunkumar and B. V. Ravishankar (2021), Strengthening of Brick Masonry Wallette by Polyester Geo-Fabric Reinforced Cementitious Matrix, Materials Today: Proceedings, Vol. 46, pp. 2392-2399.

21. International Organization for Standardization (2017), ISO/IEC 17025: General Requirements for the Competence of Testing and Calibration Laboratories, ISO, Geneva.

22. Bureau of Indian Standards (1972, Reaffirmed 2017), IS 6441 (Parts 1 to 9): Methods of Test for Autoclaved Cellular Concrete Products, BIS, New Delhi.

23. ASTM International (2020), ASTM C1693-11(2020): Standard Specification for Autoclaved Aerated Concrete (AAC), ASTM International, West Conshohocken, PA.

24. ASTM International (2020), ASTM C1585: Standard Test Method for Measurement of Rate of Absorption of Water by Hydraulic-Cement Concretes, ASTM International, West Conshohocken, PA.

25. ASTM International (2019), ASTM C177: Standard Test Method for Steady-State Heat Flux Measurements and Thermal Transmission Properties by Means of the Guarded-Hot-Plate Apparatus, ASTM International, West Conshohocken, PA.

26. Bureau of Indian Standards (1978, Reaffirmed 2021), IS 3792: Guide for Heat Insulation of Non-Industrial Buildings, BIS, New Delhi.

27. Bureau of Indian Standards (2016), IS 1893 (Part 1): Criteria for Earthquake Resistant Design of Structures - Part 1: General Provisions and Buildings, BIS, New Delhi.

28. Bureau of Indian Standards (1985, Reaffirmed 2020), IS 6041: Code of Practice for Construction of Autoclaved Cellular Concrete Block Masonry, BIS, New Delhi.

29. Bureau of Indian Standards (1987, Reaffirmed 2017), IS 1905: Code of Practice for Structural Use of Unreinforced Masonry, BIS, New Delhi.

30. Bureau of Energy Efficiency (2017), Energy Conservation Building Code 2017, Ministry of Power, Government of India, New Delhi.

31. Bureau of Indian Standards (1987, Reaffirmed 2018), IS 875 (Part 2): Code of Practice for Design Loads (Other than Earthquake) for Buildings and Structures - Part 2: Imposed Loads, BIS, New Delhi.

32. Bureau of Indian Standards (2000, Reaffirmed 2021), IS 456: Plain and Reinforced Concrete - Code of Practice, BIS, New Delhi.

33. International Organization for Standardization (2006), ISO 14044: Environmental Management - Life Cycle Assessment - Requirements and Guidelines, ISO, Geneva.

34. Ecoinvent Association (2021), Ecoinvent Database Version 3.8, Zurich, Switzerland.

DOI - 10.68648/ijte-q-2026-49-03-007

Identifying the Best Interpolation Technique for $PM_{2.5}$ Concentrations in Delhi

Sumit Suman
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Khagaria
Khagaria, Bihar
✉ suman.sumit@geckhagaria.org.in

ABSTRACT

Continuous ambient air quality monitoring stations report $PM_{2.5}$ at a small number of discrete points, whereas exposure and health-impact studies require concentrations at arbitrary locations. This study identifies the interpolation technique that predicts $PM_{2.5}$ over Delhi most accurately and, unlike comparisons that adopt software defaults, optimises the governing parameters of every technique before comparing them. Four techniques available in ArcGIS Geostatistical Analyst were examined: inverse distance weighting (IDW), spline, kriging and cokriging. For each technique three parameters were varied systematically, and leave-one-out cross-validation across 24 monitoring stations gave the root mean square prediction error for every combination, in a spatial phase (10 dates) and a temporal phase (24 stations). Cokriging gave the lowest mean error in both phases, 26.21 and 26.03 µg/m3 respectively, improving on IDW by 6.9 and 4.3 per cent; wind speed, temperature and relative humidity were selected as the optimal secondary variables in both phases. The margin over the deterministic techniques narrows sharply in the temporal phase. Prediction surfaces generated with the optimised model reproduce the two annual pollution episodes reported for Delhi.

***KEYWORDS** : $PM_{2.5}$, Spatial interpolation, Cokriging, Cross-validation, ArcGIS.*

NOTATIONS

$PM_{2.5}$	particulate matter of aerodynamic diameter < 2.5 µm (µg/m^3)
$Z_{(s0)}$	predicted value at an unsampled location
Z_i	measured value at the i^{th} sampled location
λ_i	weight assigned to the i^{th} measured value
d_i	distance between sampled and unsampled location
k	power function in IDW
γ(h)	semi-variance at lag distance h
n	number of measured values
P, M	predicted and measured concentration (µg/m^3)
S, M, T	sector type, neighbour count, technique parameter
ME, RMSE	mean and root mean square prediction error (µg/m^3)

INTRODUCTION

Air pollution causes a large share of premature mortality in urban areas; the World Health Organization attributes of the order of 0.8 million deaths a year worldwide to it. $PM_{2.5}$, particulate matter below 2.5 micrometres in aerodynamic diameter, is the most damaging of the regulated pollutants, since it remains suspended as an aerosol and penetrates deep into the respiratory tract. Delhi is among the worst affected cities in the world, and its $PM_{2.5}$ burden concentrates into two recurring episodes each dry season, the first coinciding with post-monsoon crop-residue burning in the upwind states [1].

Controlling emissions at source is the most economical strategy but is not always feasible, so there is growing demand for urban air quality assessment, which needs the concentration at any desired location so that it can be correlated with health outcomes there. In Delhi,

ambient $PM_{2.5}$ is measured at continuous ambient air quality monitoring (CAAQM) stations run by the Central Pollution Control Board (CPCB) and the Delhi Pollution Control Committee (DPCC). These are few and sit at discrete points, so concentrations elsewhere must be estimated by spatial interpolation, which rests on Tobler's first law: near things are more related than distant things [2].

Interpolation has been applied to Delhi before, but narrowly. Shukla et al. [3] mapped particulate matter across the city's supersites with ordinary kriging and IDW, reporting average errors of about 22 and 24 per cent. That study restricts itself to the primary pollutant and to a single family of techniques and, in common with the wider literature, runs each technique at the settings the software supplies. Li and Heap's review of comparative studies [4] found IDW, ordinary kriging and ordinary cokriging to be the most-used methods and identified data variation as a dominant control on their relative accuracy; yet the parameters governing how each method responds to that variation are seldom examined. Choi and Chong [5] adopt a power function of two for particulate matter on the authority of earlier work rather than testing it, and Hernández-Gress and Conchouso González [6], mapping pollutants over Mexico City, conclude that no single variogram model is universally optimal and that kriging accuracy shifts with hour and month.

Three questions therefore remain open: whether the defaults are appropriate for $PM_{2.5}$ in Delhi, whether meteorological covariates improve the prediction, and whether the ranking of techniques is stable between spatial and temporal interpolation. This study addresses all three by (i) optimising three governing parameters for each of IDW, spline, kriging and cokriging against the root mean square prediction error (RMSE); (ii) comparing the four optimised techniques in a spatial and a temporal phase using leave-one-out cross-validation; and (iii) generating prediction surfaces and time series for Delhi with the best technique.

LITERATURE REVIEW

Geostatistical interpolation is well established in meteorology. Ozturk and Kilic [7] mapped temperature and precipitation across the Aegean region of Turkey using ordinary kriging, taking geostatistical methods as a strong alternative to deterministic ones because they model the autocorrelation between measured points. Comparative studies nevertheless disagree on a single best technique. Bhunia et al. [8] compared five methods for soil organic carbon in West Bengal and found ordinary kriging superior on cross-validated RMSE, and Rabah et al. [9] ranked kriging first for water-level and chloride surfaces in the Gaza Strip. Yang et al. [10], in contrast, found IDW slightly better than the alternatives for daily rainfall over Greater Sydney, while Wei et al. [11], correcting geographically weighted regression residuals for $PM_{2.5}$ across China, found the completely regularised spline the most consistent. Simpson and Wu [12] found spline lowest in RMSE for lake sediment while kriging performed worst, and attributed that to kriging's assumptions of stationarity and normality not being met.

Read together with [4] and [6], the pattern is clear: the reported ranking is unstable, and much of that instability traces to parameters the studies did not vary. The case for meteorological covariates is separately well supported. Mangaraj et al. [13], using a dense sensor network over Delhi-NCR, found wind speed and boundary-layer height inversely correlated with surface $PM_{2.5}$ and concluded that local meteorology governs day-to-day concentrations. Cokriging is the natural way to bring such variables into an interpolator: it uses the autocorrelation of the primary variable together with its cross-correlation with up to three secondary variables, and cannot in principle do worse than kriging. Yet no study of Delhi has tested meteorological covariates inside a cokriging model.

METHODOLOGY

Study area and data

The study area is the National Capital Territory of Delhi. Data were obtained from 24 CPCB/DPCC CAAQM stations (Fig. 1) for $PM_{2.5}$ and five meteorological variables: wind speed (WS), wind direction (WD), relative humidity (RH), temperature (TEMP) and solar radiation (SR). Ten dates at ten-day intervals from 21 December 2019 to 21 March 2020 were used for the error analysis, and prediction surfaces generated at ten-day intervals from 1 October 2019 to 11 February

2020 so that both annual episodes were covered. All processing used ArcGIS 10.3 with the Geostatistical Analyst extension [14].

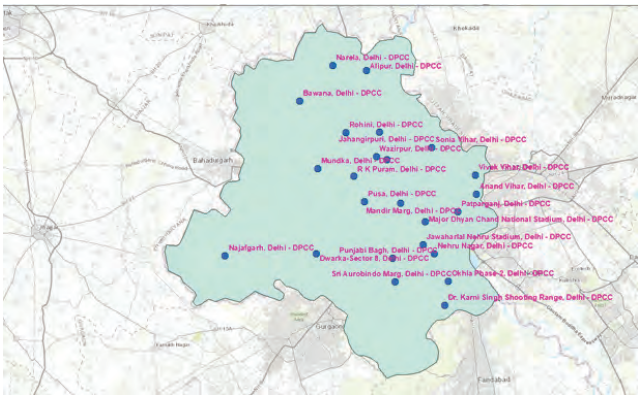


Fig. 1 The 24 CPCB/DPCC CAAQM stations used in the study, on the Delhi shapefile

Interpolation techniques

IDW predicts the value at an unsampled location as a distance-weighted mean of the measured values, the direct expression of Tobler's law:

$$Z(s_0) = \Sigma \lambda_i Z_i \,,\ \lambda_i = d_i^{-k} / \Sigma\, d_i^{-k} \qquad (1)$$

where Z is the $PM_{2.5}$ concentration, d the separation distance and k the power function; raising k reduces the influence of distant points, and k = 2 is the ArcGIS default [14]. Li and Heap [4] identify the power parameter as the main factor affecting IDW accuracy, which is why it is optimised here rather than assumed. Because the weights depend on distance alone, IDW makes no use of the structure of the data. Spline instead fits a surface through the measured points while minimising total curvature:

$$S(x, y) = T(x, y) + \Sigma \lambda_j R(r_j) \qquad (2)$$

with T and R set by the spline type: T1 (completely regularised) gives a smoother surface, T2 (with tension) one held closer to the sample range. Kriging is stochastic, predicting the value and quantifying its uncertainty; it uses the weighted-sum form of (1) but derives the weights from a semi-variogram fitted to the data, whose empirical semi-variance at lag h is

$$\gamma(h) = \tfrac{1}{2} \cdot \text{mean}\,[\,(Z_i - Z_j)^2\,] \qquad (3)$$

A stable, Gaussian, exponential or spherical model is fitted to the empirical semi-variogram, characterised by its nugget, sill and range (Fig. 2), and the weights follow from the resulting linear system. Because that system contains the covariances among the measured points themselves, kriging accounts for clustering in the network, which IDW cannot. Cokriging extends kriging by modelling $PM_{2.5}$ jointly with up to three secondary variables, so that both its autocorrelation and its cross-correlation with the covariates enter the weights.

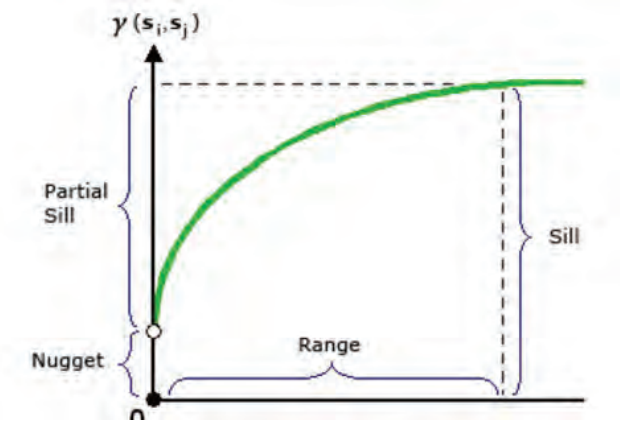


Fig. 2 Nugget, sill and range of a fitted semi-variogram

Parameters selected for optimisation

Three parameters were optimised per technique. Two are common to all four: the sector type S, controlling the shape of the search neighbourhood, and the neighbour count M, controlling its size; moving from S1 to S4 and M1 to M3 admits progressively more measured points into each estimate. The third is technique-specific: the power function k (1, 1.5, 2, 3, 4) for IDW; the spline type; the semi-variogram model for kriging; and for cokriging the choice of three secondary variables from WS, WD, RH, TEMP and SR. Levels are defined beneath Table 1.

Experimental set-up

Optimisation and comparison both rest on leave-one-out cross-validation. One parameter was varied at a time with the other two held fixed. For each setting one station was withheld, a surface was interpolated from the remaining 23, and the value at the withheld station was extracted from it. Repeating this for all 24 stations gives 24 independent prediction–measurement pairs, from which the mean error (ME) and the root mean square prediction error follow:

$$RMSE = [\ \Sigma\ (P - M)^2 / n\ ]^{½} \quad (4)$$

where P is the predicted value, M the measured value and n the number of stations. Plotting RMSE against the parameter level gives a minimum that identifies the optimum (Fig. 3). That setting was retained, the procedure repeated for the next parameter and then the next technique. The sequence was run twice: in the spatial phase the 24 stations on a single date form the validation set, repeated over ten dates; in the temporal phase the ten dates at a single station form it, repeated over 24 stations.

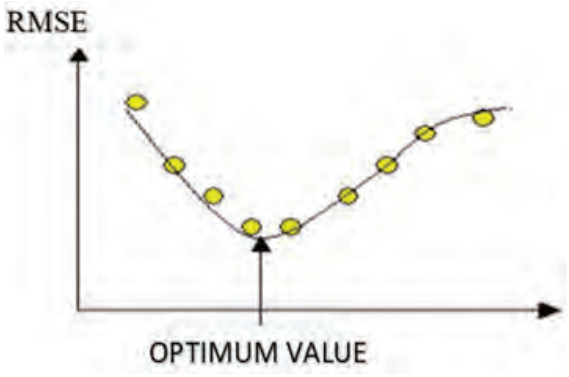


Fig. 3 Identifying the optimum level of a parameter from the RMSE minimum

RESULTS AND DISCUSSION

Optimum parameter values

Table 1 gives the parameter set that minimised RMSE for each technique in each phase. Two results stand out. First, the optimum neighbourhood is wide spatially (S3, M3) and narrow temporally (S1, M1): where the spatial field is the object of interest more stations stabilise the estimate, whereas with the location fixed, distant stations only add noise. Second, the same three secondary variables (wind speed, temperature and relative humidity) emerged as optimal for cokriging in both phases independently, which is physically reasonable and consistent with the meteorological control reported in [13]: wind governs dispersion, temperature and humidity govern boundary-layer stability and secondary aerosol formation.

The optimum power function for IDW was k = 2 spatially, confirming the default that [5] assumes, but the temporal phase gave k = 1 or k = 4 depending on the station, with no stable optimum. The default is therefore defensible for spatial work on $PM_{2.5}$ and not for temporal work. Only a parameter sweep can reveal that distinction, which is consistent with the hour- and month-dependence reported in [6].

Table 1. Optimum parameter values for each technique

Technique	Parameters optimised	Optimum (spatial)	Optimum (temporal)
IDW	Sector type, neighbours, power k	S1/S3, M3, k = 2	S1, M1, k = 1 or 4
Spline	Sector type, neighbours, spline type	S3, M3, T1	S1, M1, T1
Kriging	Sector type, neighbours, semi-variogram	S3/S4, M3, T1	S1, M1, T1
Cokriging	Secondary variables (3 of 5)	WS, TEMP, RH	WS, TEMP, RH

S1 = 1 sector; S2 = 4 sectors; S3 = 4 sectors, 45° offset; S4 = 8 sectors. M1 = 5/2, M2 = 10/6, M3 = 15/10 max/min neighbours. Spline: T1 = completely regularised, T2 = with tension. Kriging: T1 = stable, T2 = Gaussian, T3 = exponential, T4 = spherical

Spatial analysis

Table 2 lists the cross-validated RMSE of each optimised technique on the ten dates. The ranking is unambiguous: cokriging first, then kriging, then spline, with IDW last. A cokriging variant was best on nine dates out of ten, and averaged over the ten dates the best model reduced RMSE from 28.14 µg/m^3 for IDW to 26.21 µg/m^3, an improvement of 6.9 per cent. The ranking follows from what each technique uses: IDW derives its weights from distance alone; spline uses the concentrations but forces an exact fit; kriging derives weights from a fitted semi-variogram; cokriging adds the cross-correlation with meteorology. Each step up the ranking exploits additional structure.

Averaging conceals a difference between the two cokriging variants: the single-variable model has the marginally lower mean (26.21 against 26.27 µg/m^3), yet the three-variable model won on twice as many dates. The mean is dominated by 1 January 2020, the most variable date in the set, on which the simpler model did better (55.16 against 56.99 µg/m^3). Each extra covariate needs further autocorrelation parameters estimated,

and on such a day that added variance outweighs the information gained.

Table 2. Cross-validated RMSE (µg/m³) of each optimised technique for the spatial analysis

Date	IDW	Spline	Kriging	Cokriging (1 sec. var.)	Cokriging (3 sec. var.)
21 Dec 2019	31.45	31.65	31.07	29.98	**29.45**
1 Jan 2020	66.66	63.44	60.29	**55.16**	56.99
11 Jan 2020	39.56	40.17	**39.02**	**39.02**	**39.02**
21 Jan 2020	46.49	46.24	46.16	44.73	**43.85**
1 Feb 2020	22.31	21.25	20.73	19.75	**19.47**
11 Feb 2020	35.02	35.61	34.97	**34.86**	35.33
21 Feb 2020	8.23	8.03	7.97	7.93	**7.81**
1 Mar 2020	10.79	10.68	10.64	10.54	**10.24**
11 Mar 2020	5.75	5.76	5.74	5.52	**5.32**
21 Mar 2020	15.18	14.91	14.89	**14.62**	15.25
Mean RMSE	**28.14**	**27.77**	**27.15**	**26.21**	**26.27**
Mean ME	**+2.73**	**+1.84**	**+0.85**	**+0.57**	**+0.63**

Bold = lowest RMSE on that date

Temporal analysis

Table 3 gives the per-station results. Cokriging is again most accurate overall, 26.03 against 27.21 µg/m³ for IDW, but the margin narrows to 4.3 per cent and the ranking of the other three collapses: IDW, clearly worst spatially, is now second, and spline and kriging are separated by 0.02 µg/m³. Outright wins tell the same story: cokriging best at 9 stations, IDW at 8, spline at 6.

The explanation lies in what varies in each phase. Spatially, the field is $PM_{2.5}$ across Delhi on one day, and concentrations at different points in the city are broadly similar; the field is smooth and strongly autocorrelated, so a technique modelling that autocorrelation has much to exploit. Temporally, the location is fixed and the variation spans three months, over which $PM_{2.5}$ swings by more than an order of magnitude. Spatial autocorrelation is no longer the dominant structure, so cokriging's advantage is largely neutralised and the deterministic techniques close most of the gap. This is precisely the behaviour anticipated by [4], which identified data variation as a dominant impact factor and the resulting loss of accuracy as method-dependent. The choice of technique therefore matters far more for mapping than for gap-filling a station record.

Table 3. Cross-validated RMSE (µg/m³) of each optimised technique for the temporal analysis

Station	IDW	Spline	Kriging	Cokriging
Alipur	34.01	31.41	37.29	**31.36**
Anand Vihar	**12.97**	13.35	14.53	14.14
Ashok Vihar	7.19	**5.79**	7.73	7.18
Bawana	23.40	**16.89**	21.74	24.69
Dr. Karni Singh S.R.	20.44	16.38	22.82	**14.89**
Dwarka-Sector 8	**34.29**	51.60	37.95	39.00
Jahangirpuri	**20.19**	25.38	26.61	24.83
J. Nehru Stadium	17.48	17.91	**7.36**	10.18
Major Dhyan Chand Stadium	16.55	**6.89**	15.70	15.43
Mandir Marg	**20.75**	23.31	21.24	23.80
Mundka	49.41	**45.68**	49.82	51.40
Najafgarh	54.92	65.88	54.20	**52.38**
Narela	**12.14**	14.77	19.36	19.48
Nehru Nagar	58.63	55.27	56.00	**54.05**
Okhla Phase-2	24.93	**12.84**	16.92	13.33
Patparganj	17.58	20.54	17.41	**17.02**
Punjabi Bagh	**21.21**	28.28	27.11	21.75
Pusa	38.06	40.79	36.13	**32.48**
R K Puram	**46.29**	52.09	49.49	48.15
Rohini	25.57	**21.85**	23.50	24.94
Sonia Vihar	33.11	28.35	28.81	**25.90**
Sri Aurobindo Marg	34.73	38.85	36.57	**28.31**
Vivek Vihar	12.71	12.81	14.22	**10.99**
Wazirpur	**16.40**	17.27	21.11	19.13
Mean RMSE	**27.21**	**27.67**	**27.65**	**26.03**

Bold = lowest RMSE at that station.

Prediction surfaces

Surfaces were generated with the optimised cokriging model at ten-day intervals from 1 October 2019 to 11 February 2020. Predicted $PM_{2.5}$ ranged from 28.02–45.23 µg/m³ on 1 October 2019, the cleanest date, to 265.89–455.60 µg/m³ on 1 November 2019 and 256.50–493.63 µg/m³ on 1 January 2020 (Fig. 4). The clean-season surface is comparatively flat while both episode surfaces show a pronounced gradient, so the smoothness of the field is itself seasonal.

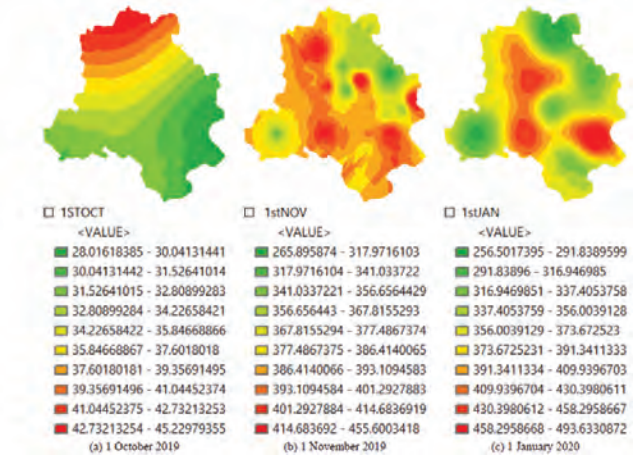
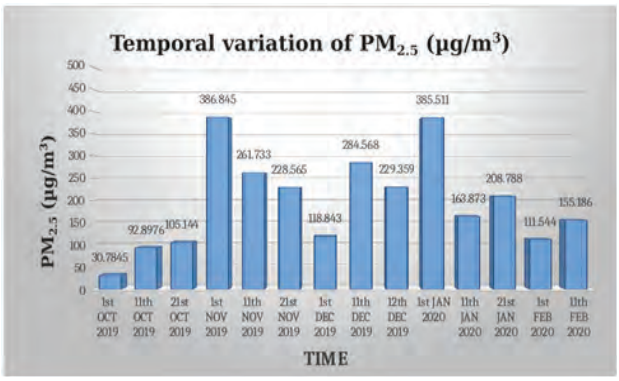


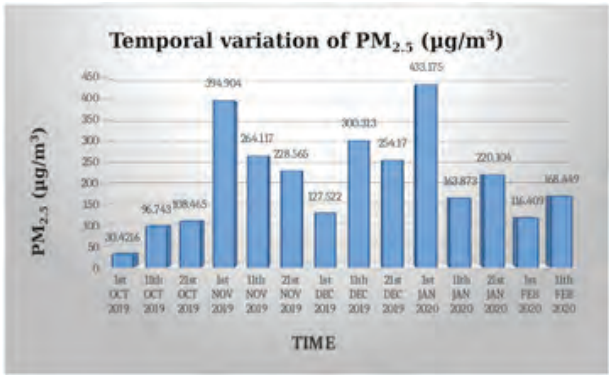
Fig. 4 Predicted $PM_{2.5}$ surfaces over Delhi from the optimised cokriging model (µg/m³); class breaks are given beneath each map

Temporal variation and episode reproduction

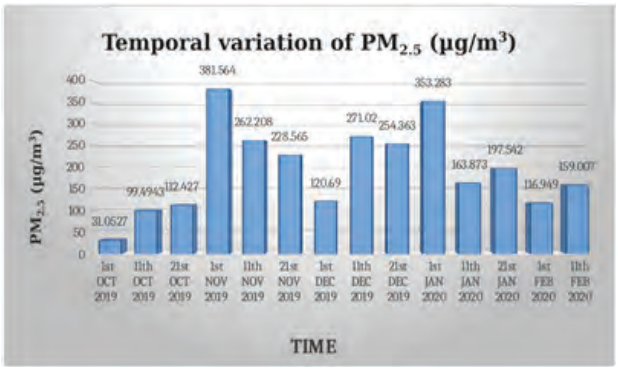
The same model was used to extract time series at three non-monitored locations (IIT Delhi, AIIMS and Connaught Place), the use case that motivates the study (Fig. 5). All three track one another closely and exceed the 24-hour national standard of 60 µg/m³ on every date after 1 October. Two peaks stand out: 1 November 2019 (382–395 µg/m³) and 1 January 2020 (353–433 µg/m³), matching the two dry-season episodes documented for the city. The first coincides with post-monsoon crop-residue burning upwind, whose contribution is itself contested: Kajino et al. [1] put it at 30-34 per cent of surface $PM_{2.5}$ over the Punjab-Haryana-Delhi region, rising to 50-56 per cent during plume transport events, whereas Mangaraj et al. [13] attribute only about 14 per cent of Delhi-NCR $PM_{2.5}$ to burning over Punjab. The second is attributed to local emissions and secondary formation under stable winter meteorology. That an independently optimised model reproduces both episodes, in the right months and with the right relative magnitudes, corroborates the predicted surfaces.



(a) IIT Delhi



(b) AIIMS



(c) Connaught Place

Fig. 5 Predicted $PM_{2.5}$ at three non-monitored locations, showing the two annual episodes

Systematic smoothing of extremes

The prediction–measurement pairs reveal a bias common to all four techniques. Fig. 6 plots predicted against measured $PM_{2.5}$ for the 24 stations; the fitted

line has a slope of 0.289, far below the 1:1 line, so high measured values are systematically under-predicted and low ones over-predicted. Every technique here estimates a value as a weighted combination of its neighbours, which can never exceed the range of those neighbours, so extremes are pulled towards the local mean.

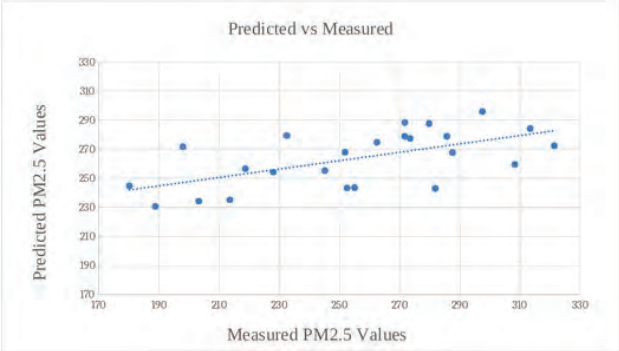


Fig. 6 Predicted versus measured $PM_{2.5}$ (µg/m³) for the 24 stations; fitted line y = 0.2887x + 189.98

Mean error across the network is nonetheless small, +0.57 µg/m³ for the best cokriging model against +2.73 µg/m³ for IDW (Table 2), because over-prediction at clean sites and under-prediction at polluted sites largely cancel. Interpolated surfaces are therefore reliable for city-wide averages but understate peak exposure at hotspots. This matters directly for the health-impact applications that motivate [3], and is a further reason to prefer cokriging, whose covariates reduce the effect.

CONCLUSIONS

Cokriging with wind speed, temperature and relative humidity as secondary variables is the most accurate of the four techniques for predicting $PM_{2.5}$ over Delhi, in both phases. The same three covariates were selected independently in each phase, so the result reflects the physics of the field rather than an artefact of one dataset. No previous study of Delhi had tested meteorological covariates inside a cokriging model.

Its advantage is conditional on what is being predicted. Spatially it is decisively better, with 6.9 per cent lower RMSE than IDW and the best result on nine of ten dates. Temporally, at a fixed location, the advantage narrows to 4.3 per cent and IDW becomes competitive. Reporting a single best interpolator without stating what is being interpolated is therefore misleading.

Optimising parameters before comparing techniques changes the comparison. The optimum search neighbourhood was wide spatially (S3, M3) and narrow temporally (S1, M1); k = 2 was confirmed for spatial interpolation of $PM_{2.5}$ but found unstable temporally. Rankings obtained at default settings measure the defaults as much as the techniques.

All four techniques smooth extremes, with a predicted-against-measured slope of 0.289, so interpolated surfaces are dependable for city-wide averages but understate hotspot peaks. The natural extension is to couple these predicted concentrations with health records at the same locations, so that the disease burden attributable to $PM_{2.5}$ can be quantified and mitigation planned accordingly.

ACKNOWLEDGEMENT

This work is carried out under the supervision of Prof. Mukesh Khare and Associate Professor, Dr. Sri Harsha Kota, Department of Civil Engineering, IIT Delhi. Their guidance throughout the work is gratefully acknowledged. The author also thanks the Central Pollution Control Board and the Delhi Pollution Control Committee for making the CAAQM data available online publicly.

REFERENCES

1. M. Kajino, K. Ishijima, J. Ching et al. (2025), “Impact of post-monsoon crop residue burning on $PM_{2.5}$ over northern India: optimizing emissions using a high-density in situ surface observation network”, Atmospheric Chemistry and Physics, Vol. 25, Pp. 7137-7160. doi: 10.5194/acp-25-7137-2025

2. Waldo Tobler (2004), “On the First Law of Geography: A Reply”, Annals of the Association of American Geographers, Vol. 94, No. 02, Pp. 304-310. doi: 10.1111/j.1467-8306.2004.09402009.x

3. Komal Shukla, Prashant Kumar, Gaurav S. Mann and Mukesh Khare (2020), “Mapping spatial distribution of particulate matter using Kriging and Inverse Distance Weighting at supersites of megacity Delhi”, Sustainable Cities and Society, Vol. 54, Article 101997. doi: 10.1016/j.scs.2019.101997

4. Jin Li and Andrew D. Heap (2008), A Review of Spatial Interpolation Methods for Environmental Scientists,

Geoscience Australia, Record 2008/23, Canberra, 137 pp. GeoCat 68229, ISBN 978-1-921498-28-2.

5. Kanghyeok Choi and Kyusoo Chong (2022), "Modified Inverse Distance Weighting Interpolation for Particulate Matter Estimation and Mapping", Atmosphere, Vol. 13, No. 05, Article 846. doi: 10.3390/atmos13050846

6. Eva Selene Hernandez-Gress and David Conchouso Gonzalez (2026), "When Does Geostatistical Interpolation Work? Monthly and Hourly Sensitivity of Ordinary Kriging for Urban Air Pollutant Mapping in Mexico City", Algorithms, Vol. 19, No. 03, Article 213. doi: 10.3390/a19030213

7. Derya Ozturk and Fatmagul Kilic (2016), "Geostatistical Approach for Spatial Interpolation of Meteorological Data", Anais da Academia Brasileira de Ciencias, Vol. 88, No. 04, Pp. 2121-2136. doi: 10.1590/0001-3765201620150103

8. Gouri Sankar Bhunia, Pravat Kumar Shit and Ramkrishna Maiti (2018), "Comparison of GIS-based interpolation methods for spatial distribution of soil organic carbon (SOC)", Journal of the Saudi Society of Agricultural Sciences, Vol. 17, No. 02, Pp. 114-126. doi: 10.1016/j.jssas.2016.02.001

9. Fahid K. J. Rabah, Said M. Ghabayen and Ali A. Salha (2011), "Effect of GIS Interpolation Techniques on the Accuracy of the Spatial Representation of Groundwater Monitoring Data in Gaza Strip", Journal of Environmental Science and Technology, Vol. 04, No. 06, Pp. 579-589. doi: 10.3923/jest.2011.579.589

10. Xihua Yang, Xiaojin Xie, De Li Liu, Fei Ji and Lin Wang (2015), "Spatial Interpolation of Daily Rainfall Data for Local Climate Impact Assessment over Greater Sydney Region", Advances in Meteorology, Vol. 2015, Article ID 563629, 12 pp. doi: 10.1155/2015/563629

11. Peng Wei, Shaodong Xie, Liming Huang et al. (2023), "Spatial interpolation of regional $PM_{2.5}$ concentrations in China during COVID-19 incorporating multivariate data", Atmospheric Pollution Research, Vol. 14, No. 03, Article 101688. doi: 10.1016/j.apr.2023.101688

12. Greg Simpson and Yi Hwa Wu (2014), "Accuracy and Effort of Interpolation and Sampling: Can GIS Help Lower Field Costs?", ISPRS International Journal of Geo-Information, Vol. 03, No. 04, Pp. 1317-1333. doi: 10.3390/ijgi3041317

13. P. Mangaraj, Y. Matsumi, T. Nakayama et al. (2025), "Weak coupling of observed surface $PM_{2.5}$ in Delhi-NCR with rice crop residue burning in Punjab and Haryana", npj Climate and Atmospheric Science, Vol. 08, Article 18. doi: 10.1038/s41612-025-00901-8

14. ArcGIS 10.3 Help: Geostatistical Analyst, Environmental Systems Research Institute (Esri), Redlands, California, 2015. Available: https://desktop.arcgis.com/en/arcmap/10.3/

DOI - 10.68648/ijte-q-2026-49-03-008

A Preference-Aligned Vision Transformer Framework for Camera-Based Safety Monitoring in Building Construction

Vishal Kumar Chaudhary
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Khagaria
Khagaria, Bihar
✉ vishal@geckhagaria.org.in

Pankaj Kumar
Assistant Professor
Dept. of Civil Engineering
BP Mandal College of Engineering, Madhepura
Madhepura, Bihar
✉ pankajkumarpk8888@gmail.com

ABSTRACT

Construction remains among the most hazardous sectors in India, where workers bear a share of global occupational hazards far exceeding their share of the labour force. Automated visual monitoring offers continuous supervision, but existing systems rely on convolutional detectors demanding large annotated datasets and specialist expertise, placing them beyond the reach of small and medium contractors. Recent multimodal language models have been evaluated only as benchmarks on static images, without a deployable architecture or any means of calibrating alerts to site conditions. This paper presents a four-layer framework integrating commodity Wi-Fi cameras with a compact hybrid convolutional-vision transformer backbone, a rule-mapping module that encodes site safety requirements as visually verifiable checks, and a preference-alignment loop in which routine confirm-or-dismiss responses from the site engineer are converted into preference pairs and used to calibrate alert severity through direct preference optimization. Three algorithms specify adaptive frame sampling, hazard severity ranking and on-site alignment. Compared with manual supervision, wearable sensing, convolutional detectors and cloud inference, the design requires no site-specific labelled dataset from the adopting contractor and is reconfigured by editing a table rather than by retraining. Limitations and field validation requirements are stated.

KEYWORDS *: Construction safety, Vision transformer, Preference alignment, Site monitoring, Personal protective equipment.*

NOTATIONS

Symbol	Description
H, W	Frame height and width
P	Patch edge length
N	Number of patch tokens
d	Embedding dimension
L	Number of encoder blocks
h	Number of attention heads
d_k, d_v	Key and value dimensions
C	Convolutional stem output channels
Δt	Sampling interval
p_{conf}	Detection confidence
w_r	Rule weight
g_z	Zone exposure factor
o	Observation record for one frame
r	Rule index
s(o,r)	Initial severity score
v(o,a)	Latent score of alert a
a_w, a_l	Preferred and rejected alert
τ	Escalation threshold
π_θ, π_r	Aligned and reference policy
β	Divergence coefficient
σ	Logistic function
N_{min}	Minimum preference-pair count

INTRODUCTION

The construction industry sustains a disproportionate share of the world's occupational injury burden. Reviewing the Indian position, Samanta and Gochhayat

report that 16.4 per cent of global occupational hazards are borne by workers in the Indian construction sector, which constitutes only 7.5 per cent of the world labour force, and identify non-use of personal protective equipment (PPE) among the leading contributory challenges [1]. Studying safety performance in the small-scale Indian construction industry, Singh and Misra found, through a multi-criteria ranking based on the dominance-based rough set approach, that the deployment of dedicated staff to verify that workers follow documented safety protocols ranks among the highest-weighted determinants of firm safety performance [2].

That finding frames the practical difficulty. Verification of compliance is valued but is labour-intensive and intermittent, and it competes for the same supervisory resource as production. Nor does training alone close the gap. In a study conducted in Hong Kong, Chung et al. found that poor worker knowledge of PPE remained the principal weakness even after mandatory basic safety training, and proposed continuous monitoring as the remedy [3].

Automated visual monitoring has therefore attracted sustained attention. Systems based on convolutional detectors have demonstrated reliable PPE compliance classification [4], simultaneous detection of fire and protective equipment [5], aerial inspection of unsafe conditions [6] and real-time alerting over consumer messaging platforms [7]. Each, however, depends on a purpose-built annotated dataset and on machine learning expertise to train and periodically retrain the model. Surveying practitioners in a developing-country setting, Yap et al. identified technological limitation, weak organisational commitment, and privacy and data security concerns among six dimensions of barriers to adopting new safety technologies [8]. For Indian construction specifically, Singh et al. applied a Grey-DEMATEL analysis and found real-time data processing, model accuracy and domain validity, and computational resource requirements to be the most influential blockades to generative artificial intelligence adoption [9].

Multimodal language models offer a route past the annotation requirement, since they interpret visual scenes from natural-language instruction without domain-specific training. Chaudhary et al. evaluated five such models on construction hazard recognition and concluded that structured prompt design offers a practical pathway for small firms; among their directions for future research they identified integrating model outputs into interactive workflows that close the loop between machine recognition and human decision-making [10]. No published architecture yet addresses that integration for continuous site operation, and none addresses a second problem that governs whether any monitoring system is used at all: the calibration of alert behaviour. Work on video anomaly detection notes that reducing false alarm rates is the dominant concern of that field [11], yet in construction monitoring the escalation threshold remains a fixed constant, insensitive to the work being performed or to the judgement of the engineer receiving the alerts.

This paper addresses both gaps. It presents the design of a framework in which commodity Wi-Fi cameras feed a compact hybrid convolutional-vision transformer backbone, site safety rules are expressed as visually verifiable checks, and the severity ranking that governs alert escalation is aligned to the preferences of the site engineer through routine feedback. Put plainly, the system watches the work through ordinary cameras and learns, from the engineer's own handling of the alerts it sends, which situations are worth flagging. The objectives are to specify the architecture at a level permitting implementation, to state the algorithms governing sampling, ranking and alignment, and to evaluate the design against existing approaches. The work is a design and feasibility study.

LITERATURE REVIEW

Vision-based compliance monitoring

Delhi et al. established the reference approach, applying transfer learning to a YOLOv3 base network to classify compliance into four categories from a dataset of 2,509 images collected at several construction sites, with the framework intended for integration into an organisation's safety workflow [4]. Kumar, Gupta and co-workers extended the scope to fire and trapped persons using YOLOv4, reporting a maximum mean average precision of 76.86 per cent on a self-constructed dataset [5]. Aerial platforms have been applied to the same problem by a separate group, who

deployed a UAV-mounted Faster R-CNN system that detects workers without helmets and issues immediate alerts, reducing the inspection workload placed on site supervisors [6]. More recently, Geetha et al. combined YOLOv8 with consumer messaging for automated notification of violations [7]. The paradigm has been extended to quality management, where Mahamivanan et al. addressed material recognition using test-time augmentation and outlier training samples to improve robustness against varying illumination [12].

Across this body of work the dependency is common: a labelled dataset must be assembled for the classes of interest, and any change to the rule set requires re-annotation and retraining. This is the mechanism by which the dataset and expertise barriers identified in [8] and [9] operate in practice.

Instrumented and sensor-based alternatives

Wearable sensing offers a different route. Arun Kumar et al. proposed a LoRa and Zigbee based architecture with helmet, shoe and glove detection motes coordinated by a local server and gateway [13]. Such systems avoid image processing but require an instrumented device per worker with attendant battery, ruggedisation and issue-and-return overheads, and they detect the presence of the instrumented article rather than its correct use. IoT-based environmental and personnel monitoring has been shown to work at the system level [3].

Transformer architectures and preference alignment

The transformer architecture [14] was adapted to images by Dosovitskiy et al., who showed that a sequence of linearly embedded fixed-size image patches processed by a standard transformer encoder attains classification performance competitive with convolutional networks [15]. Hybrid designs have since proved effective in civil engineering field conditions: Li et al. developed CrackTrNet, a lightweight U-shaped model in which a convolutional component extracts low-level features and a transformer component learns global context, reaching 97.60 per cent segmentation accuracy at a model size of 34.86 MB, between 66.12 and 87.85 per cent smaller than mainstream comparators, and deployed it in a working mobile inspection application [16]. Paulraj and Vairavasundaram combined shifted-window attention with convolutional recurrent layers for weakly supervised anomaly detection in industrial surveillance, noting explicitly that reduction of false alarm rates is the field's governing objective [11].

Alignment to human preference is the idea that a system can be taught what is wanted by being shown which of two outcomes a person prefers, rather than by being given a formula to optimise. It originates with Christiano et al., who showed that complex behaviour can be learned from such pairwise comparisons where no explicit reward function can be specified [17]. Ouyang et al. established the canonical pipeline of supervised fine-tuning, reward modelling and policy optimisation [18]. Rafailov et al. subsequently showed that the same objective can be solved with a simple classification loss on preference pairs, yielding an algorithm that is stable, computationally lightweight and free of extensive hyperparameter tuning [19]. Retzlaff et al. argue that human-in-the-loop methods are particularly advantageous precisely where the reward function is difficult or impossible to define, and set out the explanation requirements for users who are neither AI specialists nor machine learning practitioners [20]. Most directly, Zhang et al. applied supervised fine-tuning followed by direct preference optimization to multimodal disaster severity assessment. The fine-tuning stage raised accuracy from 73.64 to 78.29 per cent and improved explanation quality by approximately 25 per cent; the preference-optimization stage that followed improved interpretability and the agreement between model reasoning and human judgement, and the framework as a whole reduced high-risk misclassification [21]. The authors are careful to note that the gain attributable to preference optimization alone is modest, obtained on a preference set of 229 training instances.

That result is significant for the present work because it demonstrates, in an adjacent visual safety-assessment domain, the mechanism proposed here.

METHODOLOGY

System architecture

The proposed framework is organised as four layers, shown in Fig. 1. The separation is functional: each

layer exposes a defined interface to the next, so that a component may be substituted without redesign of the remainder. All inference is performed on site; no worker imagery leaves the premises, which addresses the privacy and data security barrier identified in [8] and removes the dependence on site connectivity implied by cloud inference.

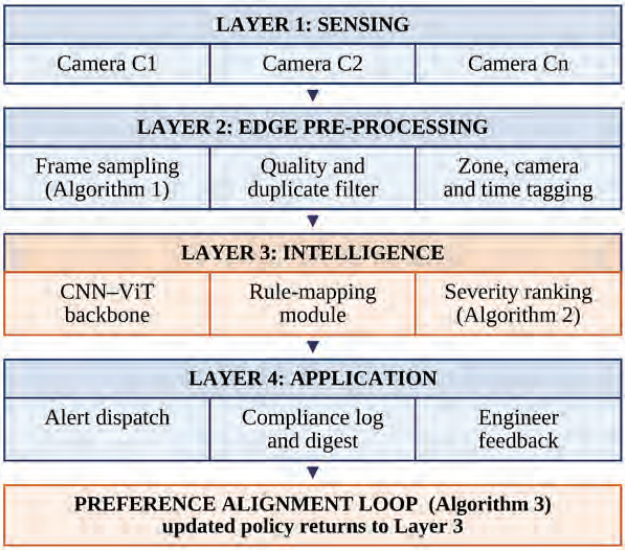


Fig. 1 Four-layer architecture of the proposed monitoring framework

Sensing layer

The sensing layer comprises commodity Wi-Fi or IP cameras of 1080p resolution supporting the real-time streaming protocol. Cameras cover defined work zones rather than the site as a whole, and are mounted at a height of 3.5 to 4.5 m with a downward tilt that keeps the working plane in the central part of the frame, where perspective distortion is least. Frames are cropped to the monitored zone and resized to the square input the backbone expects, so that the sides of the wider 16:9 sensor frame, which lie outside the zone, are discarded rather than compressed. Fig. 2 shows a typical arrangement. Two cameras give overlapping cover of the zone, so that a worker hidden from one view is usually visible in the other. Each camera carries a zone identifier and its own subset of rules, so the same installation applies different checks to a reinforcement yard and to an edge-protection zone.

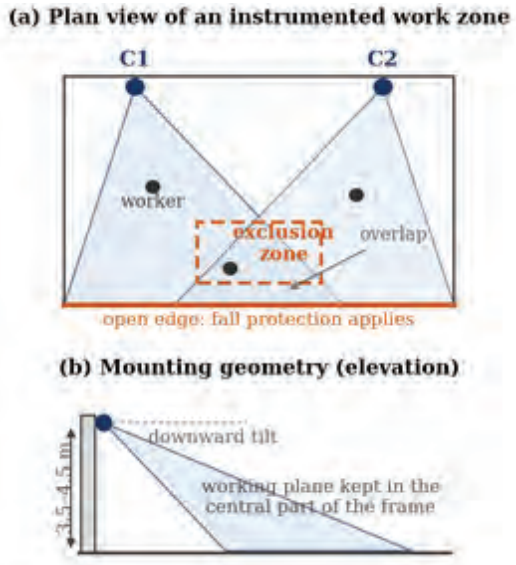


Fig. 2 Camera placement and zone coverage

Edge pre-processing layer

Continuous video is unnecessary and wasteful for the analysis performed here. The edge layer, hosted on a single-board computer, applies scheduled sampling at interval Δt together with motion-triggered capture, then discards frames that are blurred or that duplicate the preceding retained frame. Blur is assessed by the variance of the Laplacian; redundancy by the normalised absolute difference between successive frames. Algorithm 1 states the procedure. For a deployment of n cameras at $\Delta t = 300$ s over a ten-hour working day, the nominal load is 120n frames per day before filtering, which is within the capacity of modest edge hardware.

Algorithm 1: Selecting frames worth analysing

```
Input : the cameras, the sampling
        interval, and the limits for
        sharpness and change
Output: a queue of frames to analyse

1.  For each camera in turn:
2.    Capture a frame, on schedule or on
      detected movement.
3.    Measure how sharp it is. If too
      blurred, discard it and go to the
      next camera.
4.    Compare it with the last frame kept
      from that camera. If almost nothing
      has changed, discard it and go to
      the next camera.
5.    Otherwise label it with work zone,
      camera and time.
6.    Add it to the queue and keep it as
      the new reference frame.
7.  Return the queue of retained frames.
```

Intelligence layer: perception

The perception module is a compact hybrid backbone, illustrated in Fig. 3. A shallow convolutional stem extracts local features before tokenisation, following the division of labour validated for civil engineering imagery in [16]; the transformer encoder then relates every part of the frame to every other part. This matters because safety compliance is a judgement about relationships rather than about isolated objects, such as a worker standing next to an unguarded edge, or a helmet that is present but not being worn.

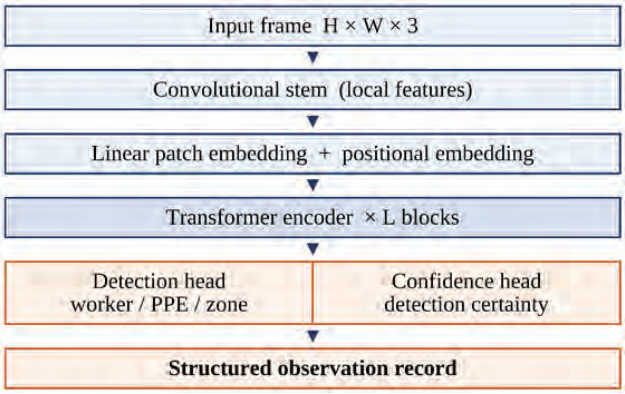


Fig. 3 Perception pipeline of the intelligence layer

Following [15], the stem output, which retains the spatial resolution H × W of the input frame, is divided into N non-overlapping patches of size P × P, where

$$N = H\,W / P^2 \qquad (1)$$

Each patch is flattened and linearly projected to dimension d, and a learnable positional embedding is added, giving the token sequence

$$z_0 = [\, x^1_p\, E\, ;\, \ldots\, ;\, x^N_p\, E\,] + E_{pos} \qquad (2)$$

The classification token of [15] is omitted, since the heads described below operate on the patch tokens themselves. Each encoder block applies multi-head self-attention, in which queries and keys have dimension d_k and values dimension d_v [14],

$$\text{Attention}(Q,K,V) = \text{softmax}(QK^T / \sqrt{d_k})\, V \qquad (3)$$

The backbone is not trained from scratch by the adopting contractor. It is initialised from transformer weights pretrained on general image collections and then fine-tuned once, by the developer, on publicly available construction safety image sets. What the framework removes is therefore not model training as such, but the requirement that each contractor label a dataset of its own before use, and repeat that whenever a rule changes. Table 1 gives the proposed configuration; the parameter budget keeps the compiled model within the size range shown to be deployable on constrained hardware in [16].

Table. 1. Proposed configuration of the perception backbone

Parameter	Value	Basis
Input H × W	640 × 640	Retains worker-scale detail across the zone
Patch size P	16	Standard granularity [15]
Tokens N	1,600	From Eq. (1)
Dimension d	384	Capacity vs. size
Blocks L	8	Reduced depth for edge
Heads h	6	d/h = 64 per head [14]
Conv. stem	3 layers	Local features [16]
Stem channels C	32	Feeds the patch embedding
Parameters	≈ 18 M	From the above configuration
Target size	< 40 MB	At half precision; cf. [16]

Intelligence layer: rule mapping and ranking

The rule-mapping module translates site safety requirements into visually verifiable conditions, each carrying a rule weight w_r reflecting the consequence of the violation. This structure is what permits reconfiguration without retraining: adding a rule adds an entry to the mapping table, not a new training cycle. Table 2 gives representative entries. Evidence that language models can be aligned to formal construction regulatory text supports extension of this table by retrieval from codified sources rather than by manual entry [22].

Table. 2. Representative rule-to-visual-check mapping

Requirement	Visually verifiable condition	w_r	Default escalation
Head protection	Helmet worn by each detected worker	0.9	Immediate
High-visibility clothing	Reflective vest on torso region	0.6	Immediate

Fall protection	Harness visible in edge zone	1.0	Immediate
Zone restriction	No worker in exclusion polygon	1.0	Immediate
Housekeeping	Access route free of material	0.4	Digest
Plant separation	No worker in plant swing radius	0.9	Immediate

For an observation o at time t comprising a detected violation of rule r with confidence p_{conf} in zone z, an initial severity score is formed as

$$s(o,r) = w_r \cdot p_{conf} \cdot g_z,\ w_r, g_z \in (0,1] \quad (4)$$

where g_z is a zone exposure factor. Both w_r and g_z lie in (0,1], so s is bounded by unity and τ may be read directly as a fraction of the maximum severity. Alerts are escalated when s(o,r) exceeds a threshold τ. In a conventional system τ is a fixed constant. This is, on the present argument, one source of the false alarm rates reported as the governing difficulty in [11]: a threshold that suppresses nuisance alerts in one work context suppresses genuine ones in another. In the proposed framework the ranking is produced by a learned policy π_θ aligned to site preference, as set out below. Algorithm 2 states the inference and ranking procedure.

Algorithm 2: Deciding which findings to report

```
Input : the frame queue, the rules for
        each zone, the ranking policy,
        the escalation threshold
Output: alerts to send, and a full log

1.  For each frame in the queue:
2.    Identify workers, protective
      equipment and zones.
3.    For each rule applying to the zone:
4.      Check whether it is broken.
5.      If so, rate the seriousness from
        the rule, the confidence of the
        analysis and the location.
6.      Adjust the rating using the
        current ranking policy.
7.      Record the finding in the log.
8.      If the rating is high enough,
        prepare an alert carrying the
        image and a short reason.
9.  Return the alerts and the log.
```

Preference alignment loop

The correct escalation for a given observation cannot be written down in advance, because it depends on the activity in progress, the stage of works and the tolerance of the site. This is the condition under which human-in-the-loop methods are held to be particularly advantageous [20]. The framework therefore treats the site engineer's routine handling of alerts as its teacher, much as a new assistant learns which matters to raise by seeing which reports are acted upon and which are waved away. Each dispatched alert carries three responses, namely confirm, dismiss and defer, and each requires a single interaction. Because a response is elicited on an alert the engineer must address in any case, the marginal burden is close to zero.

Responses are converted into preference pairs. Where one alert, a w, is confirmed and a comparable alert, a_l, dismissed, the pair is written to a buffer, $a_w > a_l$ denoting the preferred alert. Under the Bradley–Terry model, writing v(o,a) for the latent score expressing how strongly alert a merits escalation given observation o, the probability of the observed preference is

$$P(a_w > a_l \mid o) = \sigma[\, v(o,a\,w) - v(o, a_l)\,] \quad (5)$$

Following [19], the policy is obtained without fitting an explicit reward model by minimising

$$L = -E\{\log \sigma[\, \beta \log(\pi_\theta(a_w \mid o)/\pi_r(a_w \mid o)) - \beta \log(\pi_\theta(a\,l \mid o)/\pi_r(a_l \mid o))\,]\,\} \quad (6)$$

where π_r is the initial policy and β controls divergence from it. The update executes on site when the buffer reaches a minimum size N_{min}, taken as 200 pairs. Since a pair requires two comparable observations handled differently, pairs accumulate more slowly than responses; at a nominal 15 to 25 responses per working day, a first alignment cycle may be expected after roughly four to six working weeks. Retaining π_r as a reference and bounding divergence prevents the aligned policy from drifting arbitrarily far from the initial rule-derived behaviour, which is itself set from the mapping of Table 2. Fig. 4 shows the loop and Algorithm 3 states the procedure.

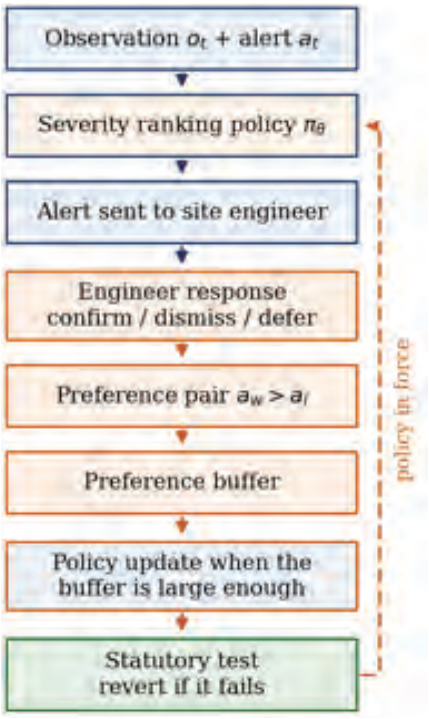


Fig. 4 Site-specific preference alignment loop

Algorithm 3: Learning the site's preferences

```
Input : the alerts sent, the reference
        policy, the preference buffer,
        the minimum buffer size
Output: an updated ranking policy

1.  For each alert sent, note whether the
    engineer confirmed, dismissed or
    deferred it. Ignore the deferred.
2.  Pair similar situations where one
    alert was confirmed and another
    dismissed; each pair records which
    the engineer preferred.
3.  Add the pairs to the buffer.
4.  Once the buffer reaches the minimum
    size:
5.    Update the policy to favour the
      alerts the engineer confirms.
6.    Test it against the statutory rules
      that must always be enforced.
7.    If the test fails, keep the
      previous policy.
8.    Discard the oldest pairs so the
      policy tracks current work.
9.  Return the policy in force.
```

Application layer

Escalated alerts are dispatched over a consumer messaging service, following the practicability demonstrated in [7], each accompanied by the annotated frame and a short rationale naming the rule and the observed condition. The explanation requirement follows [20]: a site engineer is neither an AI specialist nor a machine learning practitioner, and an alert without an intelligible basis will not be acted upon. Lower-severity observations accumulate into a daily compliance digest, and the full log supports periodic review and toolbox talks, a secondary use also identified in [10].

RESULTS AND DISCUSSION

Comparative evaluation

Table 3 evaluates the proposed design against the four alternatives available to a contractor. The comparison uses characteristics that determine adoption rather than accuracy alone, since the barriers documented in [8] and [9] are organisational and practical as much as technical.

Table. 3. Comparative analysis of construction safety monitoring approaches

Approach	Labelling by adopter	Adopter ML skill	Reconfig-uration	Alert calibration
Manual supervision	None	None	Instruction	Human judgement
Wearable sensing [13]	None	Low	Hardware change	Fixed rule
CNN detector [4,5]	≈ 2,500	High	Retraining	Fixed threshold
Cloud VLM [10]	None	Low	Prompt edit	Not addressed
Proposed framework	None by adopter	Low	Mapping-table edit	Aligned to site

Three characteristics omitted from Table 3 for space reinforce the comparison. Wearable schemes alone require a device on every worker, with the attendant issue-and-return burden; manual supervision alone gives intermittent rather than continuous coverage; and cloud inference alone requires site imagery to leave the premises.

Two differences are decisive. The first is where the labelling burden falls. The reference convolutional system had to be taught from 2,509 images hand-

labelled for that deployment [4], and every new rule calls for further labelling by the adopter. In the proposed framework that burden is borne once by the developer, using public datasets, and the rule set is thereafter extended by adding a line to a table. The second is alert calibration. No comparator adapts its alerting to the site: the deployed systems apply a fixed threshold or a fixed rule, while the benchmarking study defines no alerting mechanism at all. The surveillance literature identifies false alarm rate as its governing problem [11], and the proposed framework is the only one of the five in which escalation adapts to the judgement of the person receiving the alerts.

Expected effect of alignment

No claim of measured performance is made here. The expectation rests on the closest available evidence, and that evidence must be read precisely. In [21] the large measured gains, accuracy from 73.64 to 78.29 per cent and explanation quality by approximately 25 per cent, belong to the supervised fine-tuning stage rather than to preference optimization. What preference optimization contributed there was improved interpretability and closer agreement between model reasoning and human judgement, which its authors describe as modest and attribute to a preference set of only 229 training instances. That is nonetheless the quantity this framework needs. The objective here is not higher detection accuracy but alerting that matches the judgement of the person receiving it, which is precisely what preference optimization was found to improve. The structural correspondence is close, since both are vision-language pipelines producing a graded severity judgement aligned by human preference through a single annotation workflow, and the proposed loop accumulates preference pairs continuously rather than from a fixed set of a few hundred. The transfer is therefore a reasoned expectation, not a result.

Worker consent and acceptance

Continuous camera observation of a workforce is not a neutral act, and a monitoring system that workers regard as a disciplinary instrument will be resisted or defeated. The framework keeps imagery on site and raises alerts about conditions rather than named individuals, but neither measure substitutes for prior consultation with workers and their representatives, plain notification that the zone is monitored, and a written undertaking that the compliance log is used to correct hazards and not to penalise individuals. Where the workforce is largely unorganised, as it is for the tens of millions employed in Indian construction [1], that undertaking carries particular weight and should be treated as a condition of deployment rather than as good practice.

Limitations

Five limitations are material. First, environmental variability: illumination change, dust, rain and partial occlusion degrade visual inference, and the augmentation strategies used in [12] would form part of any implementation without eliminating the problem. Second, view stability: a construction site reconfigures as work proceeds, and formwork, scaffolding and stacked material will in time obstruct or invalidate a fixed viewpoint, so camera positions must be reviewed at each stage change, which is an operational burden the framework reduces but does not remove. Third, feedback quality: a policy aligned to one engineer inherits that engineer's inconsistencies and biases, a limitation examined generally in [23] and specifically in relation to reward misspecification and sparse feedback in [24]. No step of Algorithm 3 corrects this fully; the statutory test at step 6 catches only those biases that would breach statutory duty. Fourth, incentive alignment: the engineer who teaches the system is also the person whose site its output reflects upon, so dismissals made to reduce workload would train the system to under-report. The statutory test is the safeguard, and periodic independent audit of the compliance log should accompany deployment; alignment optimising for local preference may otherwise drift from statutory duty. Fifth, and governing, the framework is unvalidated. It is a specification supported by evidence from adjacent domains, not a measured system.

CONCLUSIONS

This paper has specified a four-layer framework for camera-based construction safety monitoring that addresses the two barriers keeping such systems out of small and medium contracting practice, namely the need for a site-specific annotated dataset and the need for machine learning expertise. By combining a compact hybrid convolutional-vision transformer backbone with a rule-mapping module, the design removes the

requirement for each contractor to assemble a purpose-built annotated dataset, and permits the rule set to be extended by editing a table rather than by retraining, answering both barriers as they are documented for developing-country construction. By converting the site engineer's routine confirm-or-dismiss handling of alerts into preference pairs and aligning the severity ranking through direct preference optimization, it addresses fixed-threshold alerting, which the present work identifies as a source of the false alarm rates reported as the governing difficulty in the surveillance literature, and it does so at close to zero marginal burden on the engineer. All inference is performed on site, which answers the privacy and data security barrier and removes dependence on connectivity.

The novelty lies in the combination: existing work either benchmarks multimodal models on static images without a deployable architecture, or deploys convolutional detectors with fixed thresholds and heavy annotation requirements.

The framework has not been validated on site data. Field implementation on an operating building construction project, measurement of detection precision and recall against expert annotation, and quantification of the change in false alarm rate before and after alignment constitute the necessary next stage. Extension of the perception layer from independent frames to short temporal sequences, and of the mapping table from safety to quality monitoring [12] and to progress monitoring [25], are identified as further directions.

REFERENCES

1. Sasmita Samanta and Jyotiranjan Gochhayat (2023), "Critique on Occupational Safety and Health in Construction Sector: An Indian Perspective", Materials Today: Proceedings, Vol. 80, Pp. 3016-3021. DOI: 10.1016/j.matpr.2021.05.707
2. Arpit Singh and Subhas C. Misra (2021), "Safety Performance & Evaluation Framework in Indian Construction Industry", Safety Science, Vol. 134, Art. 105023. DOI: 10.1016/j.ssci.2020.105023
3. William Wong Shiu Chung, Salman Tariq, Saeed Reza Mohandes and Tarek Zayed (2023), "IoT-based Application for Construction Site Safety Monitoring", International Journal of Construction Management, Vol. 23, No. 01, Pp. 58-74. DOI: 10.1080/15623599.2020.1847405
4. Venkata Santosh Kumar Delhi, R. Sankarlal and Albert Thomas (2020), "Detection of Personal Protective Equipment (PPE) Compliance on Construction Site Using Computer Vision Based Deep Learning Techniques", Frontiers in Built Environment, Vol. 06, Art. 136. DOI: 10.3389/fbuil.2020.00136
5. Saurav Kumar, Himanshu Gupta, Drishti Yadav, Irshad Ahmad Ansari and Om Prakash Verma (2022), "YOLOv4 Algorithm for the Real-time Detection of Fire and Personal Protective Equipments at Construction Sites", Multimedia Tools and Applications, Vol. 81, No. 16, Pp. 22163-22183. DOI: 10.1007/s11042-021-11280-6
6. Sourav Kumar, Mukilan Poyyamozhi, Balasubramanian Murugesan, Narayanamoorthi Rajamanickam, Roobaea Alroobaea and Waleed Nureldeen (2024), "Investigation of Unsafe Construction Site Conditions Using Deep Learning Algorithms Using Unmanned Aerial Vehicles", Sensors, Vol. 24, No. 20, Art. 6737. DOI: 10.3390/s24206737
7. G. Geetha, J. N. Swaminathan, K. Vijayalakshmi and M. Vinothkumar (2026), "Intelligent Worksite Protective Equipment Compliance Detection Using Deep Neural Networks", Journal of Engineering and Technology for Industrial Applications, Vol. 12, No. 58, Pp. 981-986. DOI: 10.5935/jetia.v12i58.3224
8. Jeffrey Boon Hui Yap, Canwin Guan Ying Lam, Martin Skitmore and Nima Talebian (2022), "Barriers to the Adoption of New Safety Technologies in Construction: A Developing Country Context", Journal of Civil Engineering and Management, Vol. 28, No. 02, Pp. 120-133. DOI: 10.3846/jcem.2022.16014
9. Arpit Singh, Anchal Patil and Shri Krishna Pandey (2026), "Unravelling the Blockades in Implementing Large Language Models in Construction Sector", International Journal of Mathematical, Engineering and Management Sciences, Vol. 11, No. 02, Pp. 991-1022. DOI: 10.33889/IJMEMS.2026.11.2.041
10. Nishi Chaudhary, S. M. Jamil Uddin, Sathvik Sharath Chandra, Anto Ovid and Alex Albert (2026), "Prompt to Protection: A Comparative Study of Multimodal LLMs in Construction Hazard Recognition", IEEE Access, Vol. 14, Pp. 70565-70580. DOI: 10.1109/ACCESS.2026.3691685
11. Shalmiya Paulraj and Subramaniyaswamy Vairavasundaram (2025), "Transformer-enabled

Weakly Supervised Abnormal Event Detection in Intelligent Video Surveillance Systems", Engineering Applications of Artificial Intelligence, Vol. 139, Art. 109496. DOI: 10.1016/j.engappai.2024.109496

12. Hadi Mahamivanan, Navid Ghassemi, Mohammad Tayarani Darbandy, Afshin Shoeibi et al. (2025), "Material Recognition for Construction Quality Monitoring Using Deep Learning Methods", Construction Innovation, Vol. 25, No. 03, Pp. 732-760. DOI: 10.1108/CI-04-2022-0074

13. G. S. Arun Kumar, Rajesh Singh, Anita Gehlot and Shaik Vaseem Akram (2022), "LoRa Enabled Real-time Monitoring of Workers in Building Construction Site", International Journal of Electrical and Electronics Research, Vol. 10, No. 01, Pp. 41-50. DOI: 10.37391/ijeer.100106

14. Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit et al. (2017), "Attention is All You Need", Advances in Neural Information Processing Systems, Vol. 30, Pp. 5998-6008. DOI: 10.48550/arXiv.1706.03762

15. Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn et al. (2021), "An Image is Worth 16×16 Words: Transformers for Image Recognition at Scale", International Conference on Learning Representations. DOI: 10.48550/arXiv.2010.11929

16. Mingchao Li, Jingyue Yuan, Qiubing Ren, Qiling Luo, Junen Fu and Zhitang Li (2024), "CNN-Transformer Hybrid Network for Concrete Dam Crack Patrol Inspection", Automation in Construction, Vol. 163, Art. 105440. DOI: 10.1016/j.autcon.2024.105440

17. Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic et al. (2017), "Deep Reinforcement Learning from Human Preferences", Advances in Neural Information Processing Systems, Vol. 30, Pp. 4299-4307. DOI: 10.48550/arXiv.1706.03741

18. Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida et al. (2022), "Training Language Models to Follow Instructions with Human Feedback", Advances in Neural Information Processing Systems, Vol. 35, Pp. 27730-27744. DOI: 10.48550/arXiv.2203.02155

19. Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon et al. (2023), "Direct Preference Optimization: Your Language Model is Secretly a Reward Model", Advances in Neural Information Processing Systems, Vol. 36. DOI: 10.48550/arXiv.2305.18290

20. Carl Orge Retzlaff, Srijita Das, Christabel Wayllace, Payam Mousavi et al. (2024), "Human-in-the-Loop Reinforcement Learning: A Survey and Position on Requirements, Challenges, and Opportunities", Journal of Artificial Intelligence Research, Vol. 79, Pp. 359-415. DOI: 10.1613/jair.1.15348

21. Yuanjun Zhang, Fuzel Ahamed Shaik, Suvojit Acharjee, Fahad Khalid and Mourad Oussalah (2026), "Towards Reliable Multimodal Disaster Severity Assessment through Preference Optimization and Explainable Vision-Language Reasoning", Reliability Engineering and System Safety, Vol. 275, Art. 112674. DOI: 10.1016/j.ress.2026.112674

22. Mohammad Aqib, Mohd Hamza, Qipei Mei and Ying Hei Chui (2025), "Fine-tuning Large Language Models and Evaluating Retrieval Methods for Improved Question Answering on Building Codes", Smart Construction, Vol. 02, No. 03. DOI: 10.55092/sc20250021

23. Stephen Casper, Xander Davies, Claudia Shi, Thomas Krendl Gilbert et al. (2023), "Open Problems and Fundamental Limitations of Reinforcement Learning from Human Feedback", Transactions on Machine Learning Research, Vol. 2023. DOI: 10.48550/arXiv.2307.15217

24. Shreyas Chaudhari, Pranjal Aggarwal, Vishvak Murahari, Tanmay Rajpurohit et al. (2025), "RLHF Deciphered: A Critical Analysis of Reinforcement Learning from Human Feedback for LLMs", ACM Computing Surveys, Vol. 58, No. 02. DOI: 10.1145/3743127

25. Aritra Pal, Jacob J. Lin, Fouad Amer, Shang-Hsien Hsieh and Mani Golparvar-Fard (2026), "Automatic Mapping of Schedule Activities and Reality Models for Tracking Construction Progress", Engineering, Construction and Architectural Management, Vol. 33, No. 03, Pp. 1932-1958. DOI: 10.1108/ECAM-08-2024-1166

DOI - 10.68648/ijte-q-2026-49-03-009

Cross-City Transfer Learning for Daily Rainfall Forecasting at Pune, India

Sumit Suman, Vishal Kumar Chaudhary
Assistant Professors
Dept. of Civil Engineering
Government Engineering College Khagaria
Khagaria, Bihar
✉ suman.sumit@geckhagaria.org.in
✉ vishal@geckhagaria.org.in

Tanmay Kumar
Assistant Professor
Dept. of Civil Engineering
Government Engineering College Samastipur
Samastipur, Bihar
✉ tanmaykumar790@gmail.com

ABSTRACT

Data-driven rainfall models normally need long local records, which many stations lack. This study asks whether a transformer pre-trained on data-rich coastal Mumbai can cut both the local data and the computation needed to forecast daily rainfall at Pune, a rain-shadow city 120 km inland. RainFormer, a transformer encoder-decoder, first trained on Pune's full record, gave a reference test RMSE of 12.522 mm. It was then pre-trained on Mumbai, and only its last decoder block and output head, 52,481 of its 194,177 parameters, were fine-tuned on 10 to 100% of Pune's record, against models trained from scratch on identical data. Transfer beat the scratch model at every share. With 10% of the local record its RMSE was within 0.43% of the reference, it equalled the reference at 50% and bettered it at 60% (12.441 mm), so full-data accuracy was reached with 40% less local data while training 73% fewer parameters. Beyond that point the advantage narrowed and the transfer RMSE rose slightly, to 12.479 mm at 100%, indicating over-fitting of the fine-tuned layers to Pune's historical regime. Cross-city transfer can therefore replace much of a station's record at a fraction of the training cost.

KEYWORDS *: Transfer learning, Transformer, Rainfall forecasting, Fine-tuning, Computational efficiency.*

NOTATIONS

Symbol	Meaning
A(S)	Transfer advantage at share S (mm)
C	Reduction in trainable parameters
D	Saving in local training data (%)
n	Number of test samples (567)
S	Percentage share of Pune's record
S_{match}	Smallest share matching Experiment A
RMSE	Root mean square error (mm)
y_i, $\hat{y}_i$	Observed and forecast rainfall (mm)

INTRODUCTION

Daily rainfall forecasts support flood preparedness and water management in Indian cities, where the southwest monsoon delivers most of the annual rain in intense spells. Extreme rain events have become markedly more frequent over central India [1], and in the 2019 monsoon heavy rain and flooding affected several districts of Maharashtra, among them were Mumbai and Pune [2].

Deep learning has improved precipitation forecasting [3], but such models learn from data and their skill depends on long, complete local records. Rain-gauge records in India vary widely in length and completeness [4], so a newly instrumented station cannot support a model trained from scratch. Transfer learning addresses this by reusing knowledge from a data-rich source for a related target with scarce data [5]. Freezing the general lower layers of a network and retraining only the task-specific upper ones [6] also reduces the number of parameters trained. In hydrology, streamflow models pre-trained on a data-rich region and fine-tuned locally have outperformed locally trained models [7], and networks trained on many catchments have been shown to learn behaviour that generalises across basins [8].

Two questions remain open for station-scale rainfall: how much local data a transferred model needs to match a model trained on the full local record, and whether the benefit persists as local data accumulate, given evidence that the advantage of pre-training shrinks as target data grow [9, 10]. This study answers both for Mumbai, on the windward coast, and Pune, 120 km inland in the rain shadow of the Western Ghats. RainFormer, a transformer encoder-decoder, was trained on Pune's full record (Experiment A) and, separately, pre-trained on Mumbai and fine-tuned on increasing shares of Pune's record, each against a model trained from scratch on the same data (Experiment B).

LITERATURE REVIEW

Deep learning for precipitation has developed largely around nowcasting, in which convolutional recurrent networks extrapolate radar fields minutes to hours ahead [3]. A radar nowcasting network named Rainformer applies window-based attention to that task [11]; it shares its name with, but is unrelated to, the station-based model used here. In India, Khan and Maity [12] predicted daily rainfall across Maharashtra, Sunil et al. [13] forecast Mumbai rainfall 6 and 12 hours ahead with a ConvLSTM2D model, and Ranjan et al. [14] combined LSTM layers with transformer encoders for eight cities using the same public dataset as this study. Each model is trained on the location it forecasts, from a long local record.

The transformer [15] replaces recurrence with self-attention and has been widely adapted to time-series forecasting; its appetite for data makes it a natural candidate for transfer learning. In hydrology, Kratzert et al. [8] showed that networks trained on many catchments learn transferable behaviour, and Ma et al. [7] transferred streamflow models from data-rich to data-sparse continents by weight initialisation and freezing. These studies concern streamflow, in which rainfall is an input rather than the quantity forecast. Transfer can also fail: the benefit of pre-training diminishes as target data increase [9], can reverse when they are plentiful [10], and fine-tuning progressively overwrites what was learned from the source [16].

Research gap. No study appears to have examined how much local data a transformer pre-trained at one Indian city needs to forecast daily rainfall at another, what computation partial fine-tuning saves, or where the benefit of transfer ends. This paper addresses that gap for the Mumbai-Pune pair.

METHODOLOGY

Study area and data

Mumbai (19.08°N, 72.88°E) lies close to sea level on the windward Konkan coast and Pune (18.52°N, 73.86°E) about 560 m above sea level, 120 km east-southeast, in the lee of the Western Ghats (Fig. 1). Monsoon flow from the Arabian Sea is forced over the Ghats, giving heavy orographic rain on the windward slopes and much lower totals inland, where rainfall at Pune also depends on the dynamical response of the flow crossing the Ghats [17].

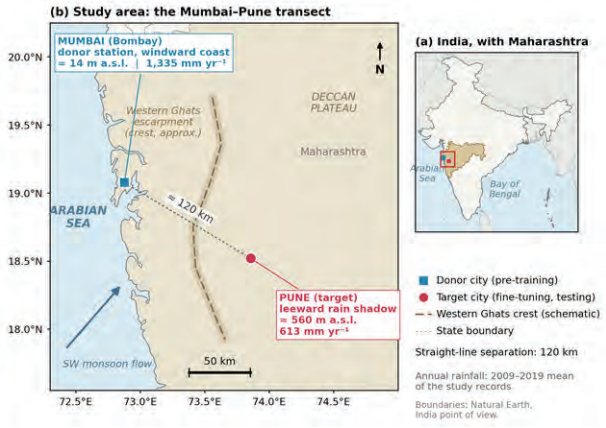


Fig. 1 Study area: (a) India, with Maharashtra highlighted (boundaries: Natural Earth, India point of view); (b) Mumbai, the donor station, and Pune, the target station, on either side of the Western Ghats.

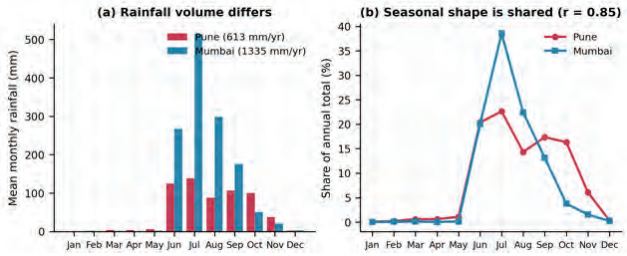


Fig. 2 Rainfall at Pune and Mumbai, 2009-2019: (a) mean monthly totals; (b) the same curves normalised by each city's annual total, showing the shared timing of the monsoon.

The data were taken from the open Kaggle collection

Historical Weather Data for Indian Cities [18], compiled from the World Weather Online historical service and used in recent Indian forecasting work [14]. The Pune and Mumbai files each hold 96,432 hourly rows of 25 fields covering 1 January 2009 to 1 January 2020; they are provider-processed city-level estimates rather than gauge observations. Mumbai receives 1,335 mm of rain a year against Pune's 613 mm, yet the monthly curves, normalised by each city's annual total, correlate at r = 0.85 (Fig. 2). In 2019 Pune received 1,787 mm, 3.6 times its 2009-2018 mean.

Data preparation

Ten uninformative columns were removed, among them clock-time strings, a snowfall column that was zero throughout and three apparent-temperature indices that duplicate air temperature. The remaining 14 variables were aggregated to daily values and expanded into 27 predictors: calendar terms, current weather, physical interactions such as the dew-point depression, and antecedent-rainfall statistics including lags and rolling means. Every history-based predictor uses only days before the forecast day, leaving 3,988 usable days per city, split chronologically into training (70%), validation (15%) and test (15%, May 2018 to January 2020) periods and scaled with each city's own training statistics. A sliding window pairs 30 days of predictors with the next three days of rainfall, giving 2,759 training and 567 test samples per city; the model predicts scaled day-to-day changes, from which forecasts are rebuilt and clipped at zero. The test period averages 3.79 mm of rain per day against 1.39 mm in training.

RainFormer model

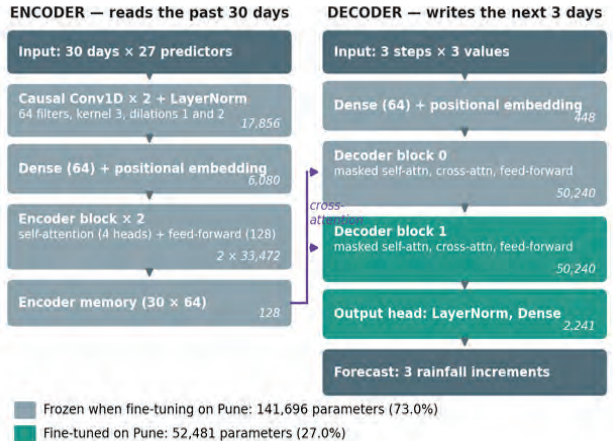


Fig. 3 RainFormer architecture, with the parameters of each block. Grey blocks are frozen and teal blocks fine-tuned in phase 2 of Experiment B; elsewhere every block is trained

RainFormer is a transformer encoder-decoder with 194,177 trainable parameters (Fig. 3). Two causal convolutions and two self-attention blocks encode the 30-day history into a memory, and two decoder blocks, each with masked self-attention, cross-attention to that memory and a feed-forward network, generate the three-day forecast through an output head. Attention follows the scaled dot-product form [15]. All models were trained with a horizon-weighted Huber loss [19] and the Adam optimiser [20], with early stopping on the validation error.

Experimental set-up

The experiments are summarised in Fig. 4 and Table 1. In Experiment A, a randomly initialised RainFormer was trained on all 2,759 Pune samples, about 7.6 years, with every parameter trainable; its test RMSE is the study's reference. Experiment B first pre-trained a new model on Mumbai, using no Pune data, and applied it to Pune's test period without fine-tuning to measure zero-shot transfer.

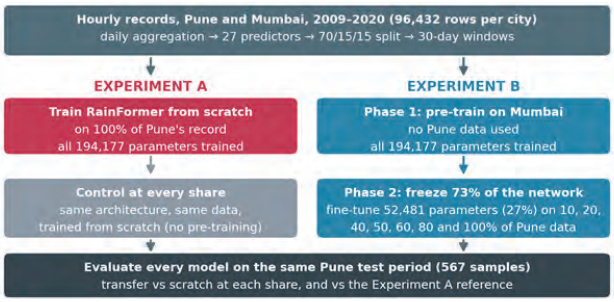


Fig. 4 Workflow: shared preprocessing, Experiment A on Pune only, and Experiment B pre-training on Mumbai and fine-tuning on shares of Pune data against a same-share scratch control.

In phase 2 the stem, both encoder blocks, the decoder projection and the first decoder block were frozen, and only the second decoder block and the output head, 52,481 parameters or 27.0% of the model, were fine-tuned on Pune data: the frozen layers hold what the cities share, the trainable layers map to Pune's smaller totals [6]. Fine-tuning was repeated on the most recent 10, 20, 40, 50, 60, 80 and 100% of Pune's samples, recent data

being used because a new station holds recent records. At each share a model was also trained from scratch on the same samples, so the difference between them isolates the contribution of pre-training. The work used TensorFlow 2.20 on one GPU with a fixed random seed, and all models were tested on the same Pune period.

Table 1. Training configuration and computational load

Setting	Experiment A (scratch)	Experiment B, phase 1 (pre-training)	Experiment B, phase 2 (fine-tuning)
Training data	Pune, 100% (2,759 samples)	Mumbai, 100% (2,759 samples)	Pune, 10-100% (276-2,759 samples)
Initial weights	Random	Random	Mumbai model of phase 1
Trainable parameters	194,177 (100%)	194,177 (100%)	52,481 (27.0%)
Frozen parameters	0	0	141,696 (73.0%)
Adam optimiser state (values)	388,354	388,354	104,962
Back-propagation through	Whole network	Whole network	Last decoder block and head
Learning rate; maximum epochs	1×10^{-3}; 90	1×10^{-3}; 70	2×10^{-4}; 50
Epochs run; training time	28; 85 s	39; 77 s	Varies with share

Performance metrics

Forecasts for the first day were evaluated over the 567 test samples by the root mean square error,

$$RMSE = \sqrt{[\ (1/n)\ \Sigma_i\ (\hat{y}_i - y_i)^2\ ]} \quad (1)$$

where ŷi and yi are forecast and observed rainfall and n = 567. Persistence (tomorrow equals today) and an always-dry forecast serve as naive baselines. Writing S for the percentage share of Pune's record used, the transfer advantage is

$$A(S) = RMSE_{scratch}(S) - RMSE_{transfer}(S) \quad (2)$$

the saving in local data is

$$D = 100 - S_{match} \quad (\%) \quad (3)$$

where S_{match} is the smallest share whose error does not exceed the reference, and the computational saving is the reduction in parameters trained,

$$C = 1 - N_{trainable} / N_{total} \quad (4)$$

RESULTS AND DISCUSSION

Reference performance

Trained on the full Pune record, RainFormer converged after 28 epochs at a test RMSE of 12.522 mm, 4.9% better than persistence (13.169 mm) and better than the always-dry forecast (12.759 mm). The margin is modest because daily point rainfall is hard to forecast and the test period, dominated by the 2019 monsoon, is 2.7 times wetter than training.

Local data needed after transfer

Applied to Pune without fine-tuning, the model pre-trained on Mumbai reached 12.728 mm, only 1.6% above the reference and better than both naive baselines: a model that has never seen Pune captures much of its rainfall behaviour, reflecting the monsoon rhythm the two cities share (Fig. 2).

Table 2 and Fig. 5 give the effect of fine-tuning. Transfer was more accurate than scratch at every share. With a 10% share, 276 samples or about nine months, transfer reached 12.576 mm, only 0.43% above the reference, whereas scratch on the same data reached 12.775 mm, 2.0% above it and worse than forecasting no rain at all. The transfer error was almost unchanged up to 40%, equalled the reference at 50% (12.525 mm) and fell below it at 60% (12.441 mm) and 80%. Full-data accuracy was therefore reached with a 60% share, a saving of D = 40%, about three years of data.

Table 2. Test RMSE of the transferred and same-share scratch models at each share of Pune data

Percentage share of Pune data	Training samples (≈ years)	Transfer RMSE (mm)	Scratch RMSE (mm)	Advantage (mm)
0% (zero-shot)	0	12.728	—	—
10%	276 (0.76)	12.576	12.775	+0.199
20%	552 (1.51)	12.574	12.721	+0.147
40%	1,104 (3.02)	12.574	12.683	+0.108

50%	1,380 (3.78)	12.525	12.635	+0.110
60%	1,655 (4.53)	12.441	12.585	+0.145
80%	2,207 (6.04)	12.442	12.514	+0.072
100%	2,759 (7.55)	12.479	12.522	+0.043

Experiment A reference 12.522 mm; persistence 13.169 mm; always-dry 12.759 mm. The scratch model at 100% is Experiment A itself.

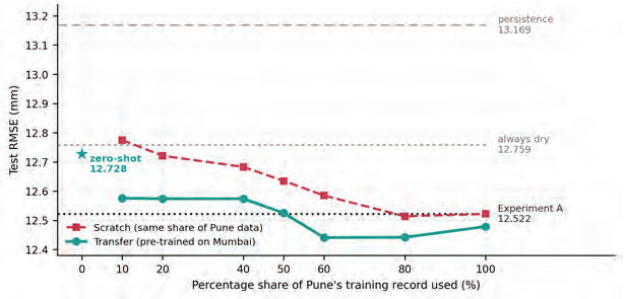


Fig. 5 Test RMSE of the transferred model and of a scratch model trained on the same share of Pune data; the dotted line is the Experiment A reference and the star the zero-shot model.

Computational benefit

Fine-tuning also reduced the computation required. Only 52,481 of the 194,177 parameters were updated, so from equation (4) the saving is C = 73.0%: gradients and weight updates covered about a quarter of the network, and back-propagation was confined to the last decoder block and the output head. The Adam optimiser keeps two moment estimates per trainable parameter [20], so it held 104,962 values instead of 388,354 (Table 1). The schedule was lighter too, at most 50 epochs against 90, and at the 60% share each epoch processed 1,655 samples rather than 2,759. The forward pass still runs through the whole network, so the saving lies in back-propagation and the weight update, not inference. Pre-training on Mumbai is a one-time cost, 77 s here, reusable for other stations.

Diminishing advantage and over-fitting

The advantage of transfer was largest where local data were scarcest, 0.199 mm at a 10% share, and fell to 0.072 mm at 80% and 0.043 mm at 100% (Table 2); averaged over the two smallest and the two largest shares it declined by two-thirds, from 0.173 to 0.058 mm. Beyond 60 to 80% the transfer error itself rose, reaching 12.479 mm with the full record, although it stayed below both the reference and the scratch model trained on the same data.

This is attributed to over-fitting of the fine-tuned layers. The model had been pre-trained on the whole Mumbai record, whose rainfall is about twice as variable as Pune's (training standard deviations 7.63 mm and 3.65 mm), so its response to heavy rain comes largely from Mumbai. Fitting the unfrozen layers to more of Pune's drier training period, 1.39 mm per day against 3.79 mm in testing, specialises them to that regime and erodes the inherited response on the heavy-rain days that dominate the RMSE, while the frozen 73% cannot improve with more data. This agrees with evidence that the benefit of pre-training diminishes as target data grow [9], can reverse when they are plentiful [10], and that fine-tuning overwrites source knowledge [16].

Earlier hydrological transfer moved streamflow models between regions [7], and Indian rainfall studies trained their models on the stations they forecast [12, 14]. This work instead quantifies, against a control trained on identical data, the local record and the computation that transfer replaces, and locates where the benefit ends.

CONCLUSION

The knowledge needed to forecast Pune's daily rainfall proves to be largely contained in the Mumbai record: what the monsoon does on the windward coast carries across the Western Ghats, and only the mapping to Pune's smaller rainfall has to be relearned locally. The contribution of this work is to measure, against a like-for-like control, how much of a station's record and how much training computation such a transfer replaces, and where it ceases to pay.

The implication is general rather than prescriptive. Where a station's record is short, a model transferred from a data-rich neighbour is the more economical option, since it approaches full-data accuracy while only about a quarter of the network is trained. As the record lengthens the benefit narrows and the two approaches converge, so further data may be better used to adapt more of the network, or to train locally; the share at which this happens will depend on the stations, the model and the period studied. The study can be

extended by repeating the experiments over several seeds, pre-training on more than one donor station, and verifying heavy-rain events against gauge observations.

REFERENCES

1. M. K. Roxy, S. Ghosh, A. Pathak et al. (2017), "A threefold rise in widespread extreme rain events over central India", Nature Communications, Vol. 8, Article 708. https://doi.org/10.1038/s41467-017-00744-9
2. A. K. Mishra, M. S. Meer and V. Nagaraju (2024), "Exploring extreme flood events of a western state of India during monsoon season of 2019 from space", MAUSAM, Vol. 75, No. 1, Pp. 245-248. https://doi.org/10.54302/mausam.v75i1.3568
3. X. Shi, Z. Chen, H. Wang et al. (2015), "Convolutional LSTM network: A machine learning approach for precipitation nowcasting", Advances in Neural Information Processing Systems, Vol. 28, Pp. 802-810.
4. D. S. Pai, L. Sridhar, M. Rajeevan et al. (2014), "Development of a new high spatial resolution (0.25° × 0.25°) long period (1901-2010) daily gridded rainfall data set over India and its comparison with existing data sets over the region", MAUSAM, Vol. 65, No. 1, Pp. 1-18.
5. S. J. Pan and Q. Yang (2010), "A survey on transfer learning", IEEE Transactions on Knowledge and Data Engineering, Vol. 22, No. 10, Pp. 1345-1359. https://doi.org/10.1109/TKDE.2009.191
6. J. Yosinski, J. Clune, Y. Bengio and H. Lipson (2014), "How transferable are features in deep neural networks?", Advances in Neural Information Processing Systems, Vol. 27, Pp. 3320-3328.
7. K. Ma, D. Feng, K. Lawson et al. (2021), "Transferring hydrologic data across continents: Leveraging data-rich regions to improve hydrologic prediction in data-sparse regions", Water Resources Research, Vol. 57, No. 5, e2020WR028600. https://doi.org/10.1029/2020WR028600
8. F. Kratzert, D. Klotz, G. Shalev et al. (2019), "Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets", Hydrology and Earth System Sciences, Vol. 23, No. 12, Pp. 5089-5110. https://doi.org/10.5194/hess-23-5089-2019
9. K. He, R. Girshick and P. Dollár (2019), "Rethinking ImageNet pre-training", Proceedings of the IEEE/CVF International Conference on Computer Vision, Pp. 4918-4927.
10. B. Zoph, G. Ghiasi, T.-Y. Lin et al. (2020), "Rethinking pre-training and self-training", Advances in Neural Information Processing Systems, Vol. 33, Pp. 3833-3845.
11. C. Bai, F. Sun, J. Zhang et al. (2022), "Rainformer: Features extraction balanced network for radar-based precipitation nowcasting", IEEE Geoscience and Remote Sensing Letters, Vol. 19, Art. no. 4023305, Pp. 1-5. https://doi.org/10.1109/LGRS.2022.3162882
12. M. I. Khan and R. Maity (2020), "Hybrid deep learning approach for multi-step-ahead daily rainfall prediction using GCM simulations", IEEE Access, Vol. 8, Pp. 52774-52784. https://doi.org/10.1109/ACCESS.2020.2980977
13. A. Sunil, A. Devda, R. Murthy and B. Deepthi (2025), "Enhanced rainfall forecasting for Mumbai using data-driven ConvLSTM2D models at fine spatial and temporal scales", Natural Hazards, Vol. 121, Pp. 11931-11956. https://doi.org/10.1007/s11069-025-07267-3
14. P. Ranjan, V. Akshay, P. R. Girijesh Kumar and M. Jagannath (2025), "Improving weather predictions accuracy with a hybrid LSTM-transformer model: A study on Indian climate", IEEE Access, Vol. 13, Pp. 169014-169025. https://doi.org/10.1109/ACCESS.2025.3614473
15. A. Vaswani, N. Shazeer, N. Parmar et al. (2017), "Attention is all you need", Advances in Neural Information Processing Systems, Vol. 30, Pp. 5998-6008.
16. J. Kirkpatrick, R. Pascanu, N. Rabinowitz et al. (2017), "Overcoming catastrophic forgetting in neural networks", Proceedings of the National Academy of Sciences, Vol. 114, No. 13, Pp. 3521-3526. https://doi.org/10.1073/pnas.1611835114
17. S. Dutta, S. I. Laskar and M. Maiti (2006), "Does the Western-Ghats play any dynamical role in the distance effect of vortex over the Bay of Bengal on the enhancement of monsoon rainfall over Pune?", MAUSAM, Vol. 57, No. 4, Pp. 567-578. https://doi.org/10.54302/mausam.v57i4.496
18. H. Soneji (n.d.), "Historical Weather Data for Indian Cities", Kaggle, https://www.kaggle.com/datasets/hiteshsoneji/historical-weather-data-for-indian-cities
19. P. J. Huber (1964), "Robust estimation of a location parameter", The Annals of Mathematical Statistics, Vol. 35, No. 1, Pp. 73-101. https://doi.org/10.1214/aoms/1177703732
20. D. P. Kingma and J. Ba (2015), "Adam: A method for stochastic optimization", Proceedings of the 3rd International Conference on Learning Representations, arXiv:1412.6980.

DOI - 10.68648/ijte-q-2026-49-03-010

The Role of Augmented and Virtual Reality in Electrical Engineering: From Education to Industry 5.0.

Malliga P
Professor
Dept. of Computer Science & Engineering
National Institute of Technical Teachers Training and Research
Chennai
✉ malliga@nitttrc.edu.in

ABSTRACT

Augmented Reality (AR) and Virtual Reality (VR) are revolutionizing various industries, and electrical engineering is no exception. This paper explores the transformative potential of AR and VR in this field. By delving into specific applications and highlighting the benefits of AR and VR in electrical engineering education, this paper underscores their pivotal role in shaping the future of the profession. The age of human-machine interaction will be emphasised in the fifth industrial revolution. Workers are free to provide clients with value-added activities based on customisation and the usage of collaborative robots. Several cutting-edge technologies and applications are being developed, which help Industry 5.0 boost production and axiomatically supply goods tailored to individual customers. Augmented Reality (AR) and Virtual Reality (VR) are two of the most significant technological developments in the last ten years. These applications are expanding quickly in many different types of settings. This paper reviews the use of VR and AR in electrical engineering education and discusses the obstacles and constraints of Industry 4.0.

KEYWORDS *: Augmented Reality, Virtual Reality, engineering education, electrical engineering, Industrial revolution.*

INTRODUCTION

Industry 4.0 refers to integrating servers, Artificial Intelligence (AI), cybernetic systems, and the Internet of Things (IoT) at the production flow level in companies and industrial processes. The idea of Industry 4.0 suggests a significant renovation of the whole production chain concerning the present industrial environment. The first part of this entails using digital technology and the Internet in existing industry. As part of the fourth industrial revolution, new digital technologies such as hardware and software are becoming widely integrated with existing ones, industrial process data and its controlling equipment are being gathered on a massive scale, and production flow has been optimised to boost output while lowering pollution. To optimise and maximise the integrity of the factory production flow, the fourth industrial revolution's concept extends beyond manufacturing procedures in the factory. It can be used to offer control over all factory processes via the Internet or AI systems.

The demands of business, technical developments, educational objectives, and wider economic, environmental and social concerns all have an impact on the electrical engineering curriculum in addition to the previously mentioned issues. The curriculum must satisfy the accreditation requirements established by its official professional organisation and address each of these criteria. There are some instances where the existing curriculum falls short of the 4.0 industry criteria. To meet the demands of Industry 4.0, future employment, and ease the industry's existing skills crisis, future graduates need not only interdisciplinary and entrepreneurial abilities but also electrical engineering skills. The electrical engineering curriculum has the potential to promote a multifaceted strategy that incorporates cutting-edge technology,

ethical issues, and sustainable practices by tackling the unique problems and demands presented by Industry 4.0. This calls for in-depth knowledge of the rapidly changing business environment, new market trends, and the moral ramifications of using cutting-edge engineering innovations.

Additionally, to guarantee that graduates have technical proficiency together with analytical thinking, problem-solving, interpersonal, and teamwork abilities, the process of developing a curriculum must consider the changing demands of businesses and employers. Students should be able to implement their understanding skills in real-world situations, get practical experience, and collaborate with industry professionals through the curriculum. The idea of AR appears to be the most pertinent at this point in the development of education. Let's take a closer look at what is meant by "AR". Technologies that allow real-time digital material to be projected into the actual environment are collectively referred to as AR. The capabilities of the programmes and devices are being employed and restrict the potential applications of AR. To present more information and improve the perception of information, hypothetical items are added to the real environment, creating AR. Depending on the user's actual location, the virtual data is fixed accordingly. The foundation for future virtualization of content is laid by its integration into the physical environment. In contrast to AR, VR uses computer technology to simulate an interactive three-dimensional environment where things have a spatial existence. AR/VR technologies are extensively employed in a variety of fields, including industrial, medical, surgical, and the military, for training purposes. VR and AR have been increasingly popular in education in recent years, offering a multitude of ways to take advantage of technology-enhanced learning. (Tan et al, 2022) VR and AR are developing ground-breaking techniques in the rapidly changing technological world that have the potential to completely transform the electrical engineering profession. There are nearly 17 research on VR applications in engineering education conducted between 2015 and 2019 and found that the best four streams were industrial, mechanical, electrical, and civil engineering (Lanzo et al, 2020).

INDUSTRY EVOLUTION IN ELECTRICAL ENGINEERING

The invention of mechanical power using steam, water, and fossil fuels in the 1780s marked the start of the First Industrial Revolution. A second similar revolution took place in the 1870s when mass production and assembly line manufacturers began to embrace electrical energy. Electronics and Information Technology (IT) were used in the 1970s to bring the notion of automation to the industrial industries as part of the Third Industrial Revolution. Cyber-physical systems, also referred to as real-time interfaces between the physical and virtual worlds, are made possible by Industry 4.0 and the fourth revolution through the usage of cloud computing and the IoT (Mourtzis et al, 2018). The first three industrial revolutions occurred over hundreds of years, while the "fourth industrial revolution" was in 2011. The emergence of digitalization is a crucial feature, driven by advancements in technologies such as AI, big data, and IoT (Mourtzis et al, 2020). Industry 5.0, which fully integrates human interaction into corporate processes and intelligent systems, is set to overtake Industry 4.0, which is still at the forefront of technological decentralization and connectivity. When people and machines collaborate, it is anticipated that the cognitive abilities of knowledgeable managers will be combined with the enhanced precision and repeatability of complete automation (European Commission, 2021).

IR 4.0 IMPLICATIONS FOR ELECTRICAL ENGINEERING EDUCATION

Many sectors are predicted to undergo substantial changes because of the Fourth Industrial Revolution (4IR), which may have an impact on the jobs and skill sets needed in the workplace. The education of electrical engineers is one area that is most affected. The 4IR places significant emphasis on technologies like cloud computing, AI, big data, machine learning and edge computing. As a result, it is evident that electrical engineering education must adapt to this new industrial landscape and will equip for this in future (K. Menon and G. Castrillón, 2019) & (Lupanda, 2023).

The Fourth Industrial Revolution has enormous and revolutionary consequences for education, especially in

the discipline of electrical engineering. The planning, development, and management of creative solutions that propel technology advancements become ever more dependent on the expertise of electrical engineering professionals as the globe grows more digitally linked and interconnected. To equip students with the abilities they need to thrive in the rapidly changing industrial world, the field of electrical engineering education needs to modify its curriculum to incorporate these cutting-edge technologies. The following are a few implications:

- Programmes in electrical engineering should be flexible enough to embrace new developments while maintaining an adequate foundation in fundamental concepts (Marwala, 2023).
- Provide students with the necessary abilities to address real-world problems, electrical engineering courses should incorporate disciplines such as AI, machine learning, cybersecurity, and data science (Bainbridge & Roco, 2016).
- Solve complicated challenges, collaborations between electrical engineers and experts from other fields including biology, computer science and materials science will increase
- In the Fourth Industrial Revolution, the capacity for critical thought and creative problem-solving will be more valuable than the recall of information. To build these crucial abilities, electrical engineering education needs to emphasize practical experiences and hands-on projects (Gleason, 2018).
- Electrical engineers will need to learn new things constantly to stay up to date with technology during their careers. Higher education institutions ought to support and enable chances for lifelong learning for industry professionals. (Naudé, 2017).
- The increasing prevalence of technology necessitates careful examination of ethical issues. To address concerns like the security of data, bias in AI, and the societal ramifications of technology advancements, ethics classes must be included in the electrical engineering curriculum.
- Environmental impact and sustainability are also becoming more widely recognised as a result of the Fourth Industrial Revolution. To solve global issues like climate change, electrical engineering programs have to place a strong emphasis on sustainable practices and green technology.

The curriculum and technology integrated teaching methods need to be administered to comply with the requirements of the Industrial revolution. The graduate attributes including lifelong learning, interdisciplinary learning, technology integration in the classroom, and adaptability develops the graduates of electrical engineering programmes ready for the rapidly changing technological innovations landscape.

AR/VR IN ELECTRICAL ENGINEERING EDUCATION

The integration of Augmented and Virtual Reality into electrical engineering curricula marks a pivotal shift from passive learning to active, immersive engagement with complex physical and theoretical systems. AR and VR helps the learners (Dixit et al,2025)::

- Visualization of Complex Concepts
- Hands-On Learning in a Virtual Environment
- Simulating Real-world Scenarios
- Enhancing Collaboration and Remote Learning
- Designing and Testing with Digital Twins

Augmented Reality in Electrical Engineering

Technologies that allow real-time digital material to be projected into the actual environment are collectively referred to as AR. The capabilities of the tools and applications restrict the potential applications of AR. The foundation for future virtualization of content is laid by its integration into the physical environment. The process of incorporating imaginary things into the actual environment to show more data and improve information perception is known as AR. Virtual symbols appearing on a pilot's helmet or the cockpit glass is a well-known application of AR. The pilot can navigate the terrain with the aid of these icons. The pilot's response time will be shortened by using these icons as he won't have to spend as much time looking over the dashboard.

The electricity sector is now using a similar solution. These are the "smart helmets" worn by power line

maintenance crew members. These employees have an application loaded on their tablet computers that provides details on the items they are serving. A label that is physically attached to the actual item is used to choose certain information. Applications that translate words from other languages using AR are a renowned example of their utilization. Smartphones frequently employ these kinds of apps. This program works based on the user pointing the camera of his smartphone at an inscription written in a foreign language. The text on the screen is then translated into the user's selected language. The language barrier is broken between individuals from different countries using AR. To accomplish well-defined educational objectives, the usage of AR in the classroom needs to be carefully controlled. Here, the wow factor turns into a diversion. However, it has been demonstrated that using AR in conjunction with conventional teaching techniques offers benefits over using eLearning and conventional teaching techniques alone (Makarova et al, 2018). This supports the Moscow State University rector's well-known theory that worse educational quality results from granting students total autonomy in selecting their online course materials. AR is currently being used in computer science, physics, chemistry, history, and astronomy classes at educational institutions (Kirgizova et al, 2020). As a result of faculty members lacking the interest to develop relevant information, AR is rarely used in higher education (Vanin et al, 2018), (Nesterov, 2017), (Nesterov, 2017, P.117), (Muzyleva et al, 2019).

Augmented Reality Working Principle

The physical world and the virtual world can be connected through AR. The Google Lens app, which is integrated with Google Photos and Google's Assistant interactive learning environment, is one method to put such a bridge into practice. It assists in scanning everything that is aimed at by the smartphone camera. This leads to the recognition of many things, including plants, animals, buildings, logos, and QR codes; moreover, phrases are spoken and translated. Figure 1 depicts the flow of information in popular AR apps.

The following three characteristics describe AR:

1. AR integrates virtual and real-world data.
2. Real-world interaction is possible with AR.
3. AR can function and be used in three-dimensional spaces.

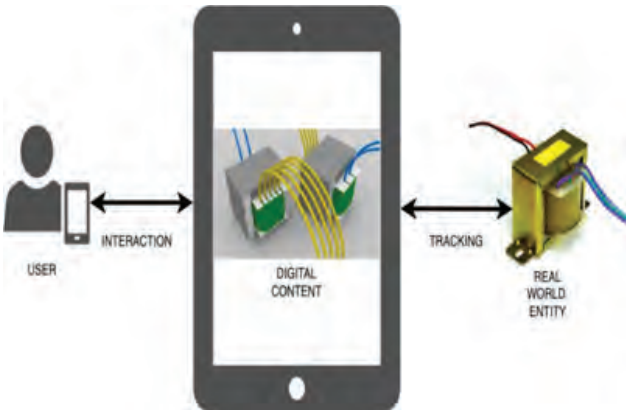


Fig. 1. Information movement in well-known AR applications

AR aims to enable real-time user interaction between the virtual and physical worlds. The virtual environment will seem real to users. AR enables the augmentation of the physical environment with digital material, such as 2D and 3D animation. There are two types of tracking capabilities that a gadget can have marker tracking and markerless tracking. An item or picture is designated as the 3D object's exit location in marker tracking (Aprinaldi et al, 2019) & (Aditya & Trisno, 2018). The terms "trackable" refer to the items or pictures. A predefined shape or several colours can be detected by the system while using markerless tracking. The primary applications in electrical engineering are depicted in Figure 2.

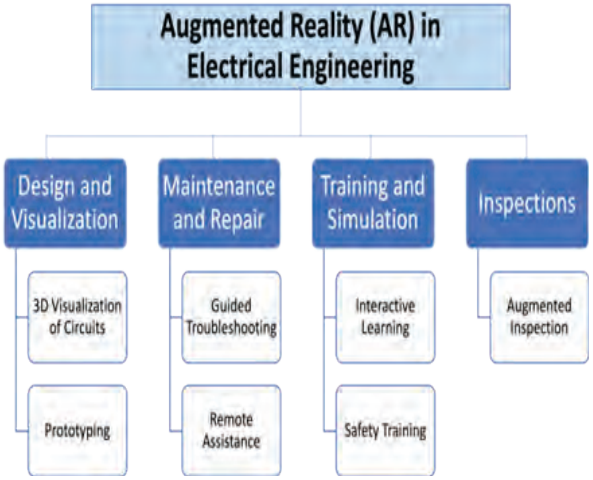


Fig. 2. The AR main applications in Electrical Engineering

Virtual Reality in Electrical Engineering

A computer-generated environment is interactive and often has a dynamically changing scene that allows users to view and move about is called VR. VR allows a user to interact with a constructed world and replicate their actual presence in it. Subsequently, these sectors employ the greatest numbers of people using VR headsets, majority of VR apps and solutions concentrate on the gaming and commercial businesses. However, VR's applications extend beyond games. VR has gained popularity and may now be used in a wide range of fields, including education, psychology, medicine, and in armed forces. It has been discovered that using information and communication technology helps students' attitudes towards studying. VR also has a significant impact on education, offering captivating and exciting methods of learning. Its graphical style, coupled with an exploratory methodology, tangible interactions, and user-friendly interfaces, can aid educators in clarifying complicated concepts.

In power engineering, VR technology is used to educate students on various power principles or to give tours of facilities that are generally inaccessible. The researchers used cloud-based architecture to create a non-immersive VR application that teaches first-year engineering students about solar array installation, photovoltaic cell technology, and solar module technology (Abichandani et al, 2019).

Virtual Reality Working Principle

VR technology is frequently used in industry to teach workers about high-voltage machinery or to train them in a particular environment. A non-immersive application created by the European business E.ON is covered in reference (Kvasznicza et al, 2019). Students studying electrical engineering may learn about high- and medium-voltage substations using this programme. It equips students with the knowledge and abilities needed to operate at E.ON. An industrial VR application for assembling and viewing a control cabinet containing power transformers and their components using multimodal channels was developed in (Shao et al, 2020) These channels provide real-time video contact between the staff and the training teacher, as well as collision detection and object capture functionality. A digital duplicate of a grid control station was created to teach staff members in a variety of scenarios, including transformer failure and back feed in electrical grids (Liegmann et al, 2022). To reduce expenses during the COVID-19 epidemic, a VR tool was created to teach staff members how to construct a transformer and its numerous components in a short time (Jiang et al, 2021). Employees were trained on the 10 kV line breaker technique for relay protection with the use of VR technology (Xu et al, 2019). An immersive VR program for switches, circuit breakers, and transformers was used to instruct electricians in Brazil (Tanaka et al, 2017). The hydropower station was created with immersive VR that allows users to practice emergency procedures, learn station maintenance, and identify defects (Liu, 2021). The number of power line accidents was reduced by teaching medium-tension live-line operators over 40 maintenance techniques using a non-immersive program developed for this purpose (Perez-Ramirez et al, 2021). An instructional instrument to assist electricians to service and maintain electrical systems was created by the authors (Górski et al, 2021). A VR operator training application for a 500kV substation was developed (Liao et al, 2017).

The 3D VR program instructs pupils on the functioning of asynchronous machines (Kvasznicza, 2017). Students may study motors, their components, and one-phase transformers with the aid of the VR laboratory simulator for electric machines known as V-LAB (Hasan et al, 2020). Students can learn about condition-based maintenance of induction motors using an immersive technology created by the authors. (Checa et al, 2021) A VR system with gamification was created to teach motor maintenance personnel for a mining firm in Brazil (Oliveira et al, 2020). Students studying mechatronics engineering were taught the components of an electric motor using immersive VR (Mahdi & Akin, 2021). One desktop VR program that instructs students on how to operate full wave rectifier circuits is called VD-Circuit (Ogbuanya & Onele, 2018). In comparison to conventional classroom teaching and learning, the authors assessed the efficacy of VR technology. The VR primary applications in electrical engineering are depicted in Figure 3.

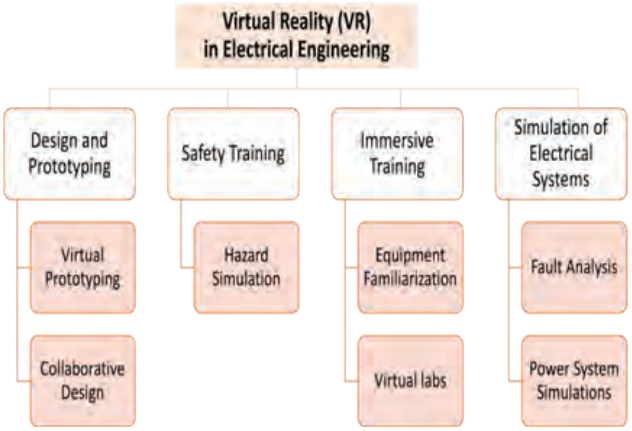


Fig. 3. VR applications in Electrical Engineering

CHALLENGES

Technological issues are the primary difficulty. These include inclement weather conditions with high light levels and smartphone incompatibility (Kaur et al, 2021), (Sendari et al, 2020) & (Sendari et al, 2021). The simultaneous recognition of several markers presented issues for AR applications (Dutta et al, 2022) & (Estrada et al, 2022). The AR application's tracking accuracy for showing the circuit analogies was low and the markers' scanning speed was decreased (Kreienbühl et al, 2020). Students reported that the AR application's reaction time was sluggish (Lucas et al, 2018). The most significant disadvantage of AR is that the programme crashes, which irritates students (Opriş et al, 2019).

The second difficulty is the level of development expertise required to make apps. This task was only referenced in two publications. The authors stated that creating a successful AR application requires prior game design skills (Tuli et al, 2022). A limitation (Martin et al, 2020) was the inability of teachers to create AR apps due to the false detection of markers. This is due to the necessity of having a basic understanding of programming languages. The authors used patterns on the marker to make them stand out from other markers to tackle the false detection problem.

The third difficulty is feeling unpleasant and uncomfortable. This problem is addressed by five VR applications and one AR application. Cognitive overload in the AR environment might result from an abundance of information (Kapp et al, 2021). Almost 45% of participants in the VR environment reported feeling unwell from cyberspace. Few people suffered from cybersickness due to the low refresh rate (measured in frames per second), which made it difficult for participants to navigate the VR environment. (Tanaka et al, 2017) The cyber illness affected 4 of the 30 participants. According to Wu et al, only a small percentage of participants felt that the HMD's weight was difficult on their eyes.

While Augmented Reality (AR) and Virtual Reality (VR) are recognized as transformative tools in pedagogy, significant barriers hinder their widespread adoption. High capital expenditure for hardware remains a primary obstacle for many academic institutions. Furthermore, the steep learning curve associated with complex system designs often results in resistance from both educators and students who lack prior technical literacy in Extended Reality (XR) environments.

Technical reliability also remains a concern; the requirement for constant hardware maintenance and software calibration increases the risk of instructional disruption. From a cognitive perspective, the highly immersive nature of these tools can occasionally become counter-productive, leading to "digital distraction" rather than enhanced retention. Within the framework of Industry 4.0, the rapid evolution of these technologies often outpaces the ability of educational curricula to adapt, necessitating costly industry-academic collaborations. Table 1 summarizes the specific challenges identified in recent literature.

Table 1: Literature on AR and VR Implementation Challenges

Methodology & Author	Key Results/ Contributions	Identified Challenges
HMD Marker T racking (Murauer et al., 2018)	Reduced error rates; scan-glove integration expedites task execution.	Increased cognitive load depending on the scanning system; visual fatigue.
Projection-Based (Doshi et al., 2017)	Reduced spot weld position dispersion (15% –52% improvement).	Precision is highly dependent on the quality of visual cues.

HMD VR (Scavarelli et al., 2021)	Explored social learning theories and advancements in VR/AR social spaces.	Mobility constraints; lack of accessibility for users with visual or motor impairments.
Multi-modal XR (Zwoliński et al., 2022)	Established XR-based learning environments using diverse hardware.	Struggles with maintaining a high "reflection of realism."
CFD & AR/ VR Integration (Solmaz & Van Gerven, 2022)	Proposed component-oriented architecture and data simulation.	Difficulties in remote data synchronization for end-user applications.
Projection-Based (Funk et al., 2017)	Improved learning curves for novices; lower NASA-TLX scores for trainees.	Significant distractions for expert workers; declining reward curves over time.

DISCUSSION ON RESEARCH GAP

To address these gaps, this review focuses specifically on the application of AR and VR within electrical engineering education, lacked a comprehensive synthesis of hardware, software, and pedagogical assessment techniques.

The proposed solution lies in the transition from the rigid automation of Industry 4.0 to the human-centric approach of Industry 5.0. While Industry 4.0 emphasizes IoT and Big Data often resulting in "over-automation" that can alienate the learner, Industry 5.0 seeks to harmonize human intuition with machine precision. By integrating AR/VR tools under the Industry 5.0 paradigm, we shift the focus toward:

- Collaborative Intelligence: Using AR/ VR to facilitate human-robot interaction rather than replacement.
- Sustainability & Resilience: Designing educational tools that are adaptable to social and environmental needs.
- Critical Thinking: Leveraging the immersive nature of AR /VR to foster high-level problem-solving rather than rote procedural tasks.

This approach mitigates the "complexity" gap identified in Industry 4.0 by prioritizing user-centric design and inclusive technological advancement.

CONCLUSION

The role of AR and VR in electrical engineering is transformative, offering educators and students innovative tools to enhance learning, visualization, and practical application of complex concepts. As these technologies continue to advance, their integration into educational curricula will be crucial in preparing the next generation of electrical engineers for a rapidly evolving and technologically driven future. The immersive nature of AR and VR not only enriches the educational experience but also equips students with the skills and insights needed to excel in the dynamic field of electrical engineering. This paper discusses the application of AR and VR in Electrical Engineering and highlights the evident need for training electrical engineering teachers in these technologies, given their potential to transform the educational landscape. Equipping educators with the tools and skills to integrate AR and VR into their teaching methodologies ensures that the next generation of electrical engineers is well-prepared for the challenges of the digital era. Embracing these technologies not only enhances the learning experience but also advances and strengthens the competitiveness of electrical engineering education.

REFERENCES

1. Abichandani, P., Mcintyre, W., Fligor, W., & Lobo, D. (2019). Solar energy education through a cloud-based desktop virtual reality system. IEEE Access, 7, 147081–147093. doi: 10.1109/ACCESS.2019.29457001
2. Aditya, F., & Trisno, B. (2018). Learning Media Based on Augmented Reality Applied on the Lesson of Electrical Network Protection System. IOP Conference Series: Materials Science and Engineering, 384(1), 012075. doi: 10.1088/1757-899X/384/1/0120756
3. Aprinaldi, A., Rahmawati, Y., & Komaro, M. (2019). Implementation of Augmented Reality (AR) android based in learning. Journal of Physics: Conference Series, 1402(7), 077045. doi: 10.1088/1742-6596/1402/7/0770454
4. Bainbridge, W. S., & Roco, M. C. (2016). Science and technology convergence: with emphasis for

nanotechnology-inspired convergence. Journal of Nanoparticle Research, 18(7). https://doi.org/10.1007/s11051-016-3520-0

5. Checa, D., Bustillo, A., Saucedo-Dorantes, J. J., Osornio-Rios, R. A., & Cruz-Albarran, I. A. (2021, October). Virtual reality-based tool applied in the teaching and training of condition-based maintenance in induction motors. In Proceedings of the IEEE 8th International Conference on e-Learning in Industrial Electronics (ICELIE) (pp. 1-6). Toronto, ON, Canada. doi:10.1109/ICELIE53900.2021.9765543

6. Di Lanzo, J. A., Valentine, A., Sohel, F., Yapp, A. Y. T., Muparadzi, K. C., & Abdelmalek, M. (2020). A review of the uses of virtual reality in engineering education. Computer Applications in Engineering Education, 28(3), 748–763. https://doi.org/10.1002/cae.22243

7. Dixit, P., & Waghmode, D. S. (2025). Virtual electrical labs powered by AR. International Journal of Innovative Research in Electrical, Electronics, Instrumentation and Control Engineering, 13(6). https://doi.org/10.17148/IJIREEICE.2025.13630

8. Doshi, A., Smith, R. T., Thomas, B. H., et al. (2017). Use of projector-based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. The International Journal of Advanced Manufacturing Technology, 89, 1279-1293.

9. Dutta, R., Mantri, A., & Singh, G. (2022). Evaluating system usability of mobile augmented reality application for teaching Karnaugh-maps. Smart Learning Environments, 9(1), 6. , doi: 10.1186/s40561-022-00189-8.

10. Estrada, J., Paheding, S., Yang, X., & Niyaz, Q. (2022). Deep-learning incorporated augmented reality application for engineering lab training. Applied Sciences, 12(10), 5159. doi:10.3390/app1210515

11. European Commission, Directorate-General for Research and Innovation, Breque, M., De Nul, L., Petridis, A. (2021). Industry 5.0: towards a sustainable, human-centric and resilient European industry, Publications Office of the European Union. https://data.europa.eu/doi/10.2777/308407

12. Funk, M., Bächler, A., Bächler, L., et al. (2017). Working with augmented reality? A long-term analysis of in-situ instructions at the assembly workplace. In Proceedings of the 10th International Conference on Pervasive Technologies Related to Assistive Environments (pp. 222-229).

13. Gleason N.W (2018) , "Higher Education in the Era of the Fourth Industrial Revolution," High. Educ. Era Fourth Ind. Revolut., pp. 1–229.

14. Górski, F., Grajewski, D., Bun, P., & Zawadzki, P. (2021). Study of interaction methods in virtual electrician training. IEEE Access, 9, 118242-118252. doi:10.1109/ACCESS.2021.3106826

15. Gottfried, M. A., & Plasman, J. S. (2018). Linking the timing of career and technical education coursetaking with high school dropout and college-going behavior. American Educational Research Journal, 55(2), 325–361

16. Hasan, B., Al-Quorashy, Y., Al-Mousa, S., Al-Sahhaf, Y., & El-Abd, M. (2020, April). V-LAB—The virtual electric machines laboratory. In Proceedings of the IEEE Global Engineering Education Conference (EDUCON) (pp. 72-77). Porto, Portugal. doi:10.1109/EDUCON45650.2020.9125349

17. Jiang, Z., Yang, Y., Yuan, Q., Leng, P., Liu, Y., & Pan, Z. (2021). Virtual reality training environment for electric systems. In Proc. IEEE 7th Int. Conf. Virtual Reality (ICVR) (pp. 314–318). Foshan, China. doi: 10.1109/ICVR51878.2021.9483825

18. Kapp, S., Lauer, F., Beil, F., Rheinländer, C. C., Wehn, N., & Kuhn, J. (2021). Smart sensors for augmented electrical experiments. Sensors, 22(1), 256. doi:10.3390/s22010256

19. Kaur, D. P., Mantri, A., & Horan, B. (2021). A framework utilizing augmented reality to enhance the teaching-learning experience of linear control systems. IETE Journal of Research, 67(2), 155-164. doi:10.1080/03772063.2018.1532822

20. Kirgizova, Ye. V., Shakirov, I. Sh., Zakharova, T. V., & Rubtsov, A. V. (2015). "Dopolnennaya Real'nost'": Innovatsionnaya Tekhnologiya Organizatsii Obrazovatel'nogo Protsessa Po Informatike ("Augmented Reality": Innovative Technology Of Organization Of The Educational Process In Informatics). Sovremennyye problemy nauki i obrazovaniya, 2-2. Retrieved from here (accessed: August 9, 2020).

21. Kreienbühl, T., Wetzel, R., Burgess, N., Schmid, A. M., & Brovelli, D. (2020, November). AR circuit constructor: Combining electricity building blocks and augmented reality for analogy-driven learning and experimentation. In Proceedings of the IEEE International Symposium on Mixed and Augmented

Reality Adjunct (ISMAR-Adjunct) (pp. 13-18). Recife, Brazil. doi:10.1109/ISMARAdjunct51615.2020.00019

22. Kvasznicza, Z. (2017, September). Teaching electrical machines in a 3D virtual space. In Proceedings of the 8th IEEE International Conference on Cognitive Infocommunications (CogInfoCom) (pp. 385-388). Debrecen, Hungary. https://doi.org/10.1109/coginfocom.2017.8268276

23. Kvasznicza, Z., Kovács, J., Maza, G., & Péli, M. (2019). VR based dual education at EON the win-win-win situation for companies, graduates and universities. In Proc. 10th IEEE Int. Conf. Cognit. Infocommunications (CogInfoCom) (pp. 479–482). Naples, Italy. doi: 10.1109/CogInfoCom47531.2019.90899612

24. Liao, X., Niu, J., Wang, H., & Du, B. (2017, October). Research on virtual reality simulation training system of the substation. In Proceedings of the International Conference on Virtual Reality and Visualization (ICVRV) (pp. 413-414). Zhengzhou, China. doi:10.1109/ICVRV.2017.00101

25. Liegmann, F., Schorge, L., Kelker, M., & Haubrock, J. (2022). Virtual reality grid control station for learning the operational management of real electrical grids on emergency situation. In Proc. IEEE German Educ. Conf. (GeCon) (pp. 1–6). Berlin, Germany. doi: 10.1109/GeCon55699.2022.99427534

26. Liu, W. (2021, December). Research and application of virtual reality technology in hydropower station production. In Proceedings of the IEEE Conference on Telecommunications, Optics and Computer Science (TOCS) (pp. 110-113). Shenyang, China. doi:10.1109/TOCS53301.2021.9688855

27. Lucas, P., Vaca, D., Domínguez, F., & Ochoa, X. (2018, July). Virtual circuits: An augmented reality circuit simulator for engineering students. In Proceedings of the IEEE 18th International Conference on Advanced Learning Technologies (ICALT) (pp. 380-384). Mumbai, India. doi:10.1109/ICALT.2018.00097

28. Lupanda,(2020) "The Impact of 4IR Technology Advancement on the Engineering Sector,". https://www.polity.org.za/article/the-impact-of-4ir-technology advancement-on-the-engineering-sector-2018-11-

29. Mahdi, S., & Akin, E. (2021). Virtual reality as a tool for electrical machines assembling and testing. International Journal of Science and Business, 5(6), 59-66. doi:10.5281/zenodo.4709807

30. Makarova, I., Shubenkova, K., & Pashkevich, A. (2018). Blended learning technologies in the automotive industry specialists' training. Proceedings - 32nd IEEE International Conference on Advanced Information Networking and Applications Workshops (WAINA 2018), 2018-January, 319–3241

31. Martin, S., Parra, G., Cubillo, J., Quintana, B., Gil, R., Perez, C., & Castro, M. (2020, July). Design of an augmented reality system for immersive learning of digital electronic. In Proceedings of the 14th Technology Applied to Electronics Teaching Conference (TAEE) (pp. 1-6). Porto, Portugal. doi:10.1109/TAEE46915.2020.9163704

32. Marwala T, (2020) "Education in the fourth industrial revolution - University of Johannesburg,". https://www.uj.ac.za/news/education-in-the-fourth industrial-revolution/

33. Menon, K., & Castrillón, G. (2019). Reimagining curricula for the fourth industrial revolution. The Independent Journal of Teaching and Learning, 14(2), 6-19.

34. Mourtzis D., (2020), Simulation in the Design and Operation of Manufacturing Systems: State of the Art and New Trends, International Journal of Production Research, 58/7, 1927–1949, DOI: https://doi.org/10.1080/- 00207543.2019.1636321

35. Murauer, N., Pflanz, N., & von Hassel, C. (2018). Comparison of scan-mechanisms in augmented reality-supported order picking processes. CEUR Workshop Proceedings, 2082, 69-76.

36. Muzyleva, I., Yazykova, L., & Penkov, V. (2019). Digitalization of the University - What is It? In 1st International Conference on Control Systems, Mathematical Modelling, Automation and Energy Efficiency (SUMMA) (pp. 335–337). Lipetsk, Russia. DOI:10.1109/SUMMA48161.2019.89475113

37. Naudé W. (2017), "Entrepreneurship, Education and the Fourth Industrial Revolution in Africa," IZA Discuss. Pap. No. 10855, no. 10855, pp. 1–25,

38. Nesterov, A. S., Kholodilin, I. Yu., Shishkov, A., & Vanin, P. (2017). Augmented reality in engineering education: Opportunities and advantages. Communications - Scientific Letters of the University of Zilina, 19(4), 117–1202

39. Oliveira, T. R. d., Martinelli, T. F., Bello, B. P., Batista, J. D., Silva, M. M. d., Rodrigues, B. B., Spinassé, R. A. N., Andreão, R. V., Mestria, M., & Schmidt, M. Q. (2020, November). Virtual reality system for industrial

motor maintenance training. In Proceedings of the 22nd Symposium on Virtual and Augmented Reality (SVR) (pp. 119-128). Porto de Galinhas, Brazil. doi:10.1109/SVR51698.2020.00031

40. Ogbuanya, T. C., & Onele, N. O. (2018). Investigating the effectiveness of desktop virtual reality for teaching and learning of electrical/electronics technology in universities. Computers in the Schools, 35(3), 226-248. doi:10.1080/07380569.2018.1492283

41. Orang, S. (2016). Aplikasi Augmented Reality Sebagai Media Simulasi Ikatan Kimia Berbasis Android Menggunakan Metode FAST CORNER DETECTION SKRIPSI Oleh: Mohammad Syahrofi Irsyad. Retrieved from here.

42. Opriş, I., Costinaş, S., Ionescu, C. S., & Nistoran, D. E. G. (2019, March). Experiencing augmented reality in power engineering education. In Proceedings of the 11th International Symposium on Advanced Topics in Electrical Engineering (ATEE) (pp. 1-6). Bucharest, Romania. doi:10.1109/ATEE.2019.8724992

43. Perez-Ramirez, M., Arroyo-Figueroa, G., & Ayala, A. (2021). The use of a virtual reality training system to improve technical skill in the maintenance of live-line power distribution networks. Interactive Learning Environments, 29(4), 527-544. doi:10.1080/10494820.2019.1587636

44. Scavarelli, A., Arya, A., & Teather, R. J. (2021). Virtual reality and augmented reality in social learning spaces: A literature review. Virtual Reality, 25(1), 257-277.

45. Sendari, S., Anggreani, D., Jiono, M., Nurhandayani, A., & Suardi, C. (2020, September). Augmented reality performance in detecting hardware components using marker-based tracking method. In Proceedings of the 4th International Conference on Vocational Education and Training (ICOVET) (pp. 1-5). Malang, Indonesia. doi:10.1109/ICOVET50258.2020.9229895

46. Sendari, S., Wibawanto, S., Jasmine, J., Jiono, M., Puspitasari, P., Diantoro, M., & Nur, H. (2021, October). Integrating robo-PEM with AR application for introducing fuel cell implementation. In Proceedings of the 7th International Conference on Electrical, Electronics, and Information Engineering (ICEEIE) (pp. 145-150). Malang, Indonesia. doi:10.1109/ICEEIE52663.2021.9616683

47. Shao, X., Feng, X., Yu, Y., Wu, Z., & Mei, P. (2020). A natural interaction method of multi-sensory channels for virtual assembly system of power transformer control cabinet. IEEE Access, 8, 54699–54709. doi:10.1109/access.2020.29815393

48. Solmaz, S., & Van Gerven, T. (2022). Automated integration of extract-based CFD results with AR/VR in engineering education for practitioners. Multimedia Tools and Applications, 81(11), 14869-14891./

49. Tan, Y., Xu, W., Li, S., & Chen, K. (2022). Augmented and virtual reality (AR/VR) for education and training in the AEC industry: A systematic review of research and applications. Buildings, 12(10), 1529.

50. Tanaka, E. H., Paludo, J. A., Bacchetti, R., Gadbem, E. V., Domingues, L. R., Cordeiro, C. S., Giraldi, O., Gallo, G. A., da Silva, A. M., & Cascone, M. H. (2017, March). Immersive virtual training for substation electricians. In Proceedings of the IEEE Virtual Reality (VR) (pp. 451-452). Los Angeles, CA, USA. doi:10.1109/VR.2017.7892374

51. Tuli, N., Singh, G., Mantri, A., & Sharma, S. (2022). Augmented reality learning environment to aid engineering students in performing practical laboratory experiments in electronics engineering. Smart Learning Environments, 9(1), 26. doi:10.1186/s40561-022-00207-9

52. Vanin, P., Nesterov, A. S., & Kholodilin, I. (2018). Integration of IIoT and AR Technologies to Educational Process Through Laboratory Complex. Proceedings – 2018 Global Smart Industry Conference, GloSIC 2018

53. Ward S. (2017.), "The Impact of Shared Leadership on a Career and Technical Education Campus," vol. 13, no. 1

54. Wu, B., Hu, Y., & Wang, M. (2020). How do head-mounted displays and planning strategy influence problem-solving-based learning in introductory electrical circuit design? Educational Technology & Society, 23(3), 40-52. Retrieved from https://www.jstor.org/stable/26926425

55. Xu, X.,Shen, L., Li, S., & Wang, S. (2019, August). Research on 10 kV line breaker check training system based on virtual reality. In Proceedings of the International Conference on Smart Grid and Electrical Automation (ICSGEA) (pp. 18-21). Xiangtan, China. doi:10.1109/ICSGEA.2019.00013.

DOI - 10.68648/ijte-q-2026-49-03-011

X-DRIFT: A Dual-Stage Mutual Information Pruning Framework for Explanation Stability in Edge-IoT Intrusion Detection Systems

Minakshi Panwar
Lecturer
Computer Science
TTC and LRDC
Jodhpur, Rajasthan
✉ minakshipanwar@gmail.com

ABSTRACT

The use of Explainable Artificial Intelligence (XAI) in Internet of Things (IoT) network Intrusion Detection Systems (IDS) allows security analysts to gain insights into how and why network traffic is identified as malicious. However, previous XAI-IDS publications assume noise-free, static lab environments. In reality, this is rarely true because of the presence of physical sensors with thermal noise, quantization errors, and environmental drift. This paper introduces a lightweight auditing and defense framework that systematically quantifies and mitigates Explanation Drift, dubbed X-DRIFT. "Explanation Drift" refers to the degradation in the quality and consistency of the explanation. Three novel auditing metrics are introduced: The Explanation Instability Metric (EIM), the Top-K Feature Retention Rate (TKFRK), and the Efficiency-Stability Tradeoff Score (ESTS). We conducted experimental evaluations on a benchmark suite of features from UNSW-NB15 dataset (N=257,673) across five noise levels. Results demonstrate that our proposed dual-stage LightGBM model with Mutual Information feature pruning achieves superior explanation stability (EIM=0.2980 at σ=0.05) and retains 60% of Top-5 features while sustaining microsecond inference ($\approx$ 1222 μs). This establishes the first empirical benchmark for XAI reliability in noisy IoT environments, outperforming standard baselines.

KEYWORDS *: Intrusion Detection System, Explainable AI, Internet of Things, Explanation Drift, LightGBM.*

INTRODUCTION

The Internet of Things (IoT) has grown to include an estimated 15 billion devices in 2024 [5, 9, 22] in areas such as: smart grids, intelligent transportation systems, monitoring systems for healthcare, and industrial control systems. This rapid growth has also increased the number of points to attack IoT networks, and so the need for sophisticated Intrusion Detection Systems (IDS) to secure them has grown in importance [17, 23, 28]. It is observed that machine learning (ML) and deep learning (DL) models are capable of providing high classification accuracy on the popular network intrusion datasets like TON_IoT [1, 5, 11] and UNSW-NB15 [17, 20, 28] with F1-scores of greater than 95% . These models, however, are overwhelmingly opaque "black boxes" that offer little detail into how they are making security decisions [8, 9, 31]. This opacity makes deployment in regulated environments, where accountability and auditability are required, difficult. As a result, there is a growing body of research into the incorporation of Explainable Artificial Intelligence (XAI), which is most commonly known as explainable artificial intelligence (XAI) tools, into IDS architectures [6, 17, 18, 23, 28] such as SHapley Additive exPlanations (SHAP) [7, 19] and LIME (Local Interpretable Model-agnostic Explanations)

Problem Statement

An underlying premise in the current XAI-IDS literature is that evaluations of feature importance explanations only take place in clean, noise-free laboratory settings [6, 10, 17, 23, 28]. This is not true in the real world

of IoT deployments, however, because there are three types of disturbances that are certain:

- Thermal noise Measured network flow features are perturbed by thermal noise from low-cost edge sensors, such as Raspberry Pi, ESP32 and Arduino. The quantization noise power for an n-bit ADC is [14,21] equals to $\sigma q2 = \Delta 2/12$, where Δ is the quantization step size .
- Precision quantization (INT8/INT16 representation) on microcontrollers makes the values of the features different, because the value is truncated. If x is quantized to $x\hat{} = \text{round}(x/s)\cdot s$ with scale factor s, the truncation error $|x-x\hat{}| \leq s/2$ is non-uniformly propagated over the features [3, 4].
- Environmental drift In the real world, the distributions of features are drifting over time because the network conditions, the number of devices, and the traffic patterns are changing, thus resulting in concept drift in the distribution P(X,Y) [13, 15, 26].

Consider a concrete example. Under clean conditions, source bytes (sbytes) may be the dominant feature for detecting Denial of Service (DoS) attacks. Sensor noise can perturb the attribution scores enough that an unrelated feature source time-to-live (sttl) rises to the top of the importance ranking. The explanation degrades, and a security analyst acting on that explanation draws the wrong conclusion.

Contributions

This paper makes the following contributions:

Explanation Instability Metric (EIM): It is a formal bounded [0, 1] metric to quantify the degradation of SHAP feature importance rankings when sensors are perturbed by the Spearman rank correlation, with proven monotonicity properties.

Top-K Feature Retention Rate (TKFR): A metric that quantifies the extent to which the most security critical features are retained as the amount of noise grows, directly capturing loss of operational interpretability.

Efficiency-Stability Tradeoff Score (ESTS): A unified score that integrates three key metrics that reflects the viability of an edge deployment in real-world applications: classification accuracy, explanation stability, and inference latency.

Dual-Stage LightGBM with MI Pruning: Proposed lightweight model architecture for MI based feature pruning and Bayesian-optimized LightGBM for deployment on the edge in an ex- planation stable manner.

Empirical Benchmark: Here I conducted the first systematic analysis of the stability of XAI explanations across five noise levels, for the full UNSW-NB15 dataset (N = 257,673) to compare the proposed model with five baselines: Decision Tree, Random Forest, XGBoost, MLP and SVM.

LITERATURE REVIEW

Literature on the application of ML to ID in IoT can be summarized on four key dimensions, each of which is highlighting a different research concern.

Key Dimension 1: Deep and Hybrid ML for IoT IDS

The use of deep learning architectures for high performance intrusion detection has been widely researched. Keshk et al. [17] came up with a framework of explainable deep learning that was tested on UNSW-NB15 and TON IoT, showing detection rates higher than 98% achieved through layer-wise feature extraction. In a recent study, Sharma et al. [28] used deep networks to complement gradient-boosted ensemble methods and achieved an F1-score of 0.978 on UNSW- NB15 using explanations from the SHAP-based post- hoc model. Ogunseyi and Thiyagarajan [22] proposed an IDS based on LSTM network for the Firefly optimized traffic patterns with competitive detection rate for the IoT traffic patterns. Sadhwani et al. [25] have shown that multi-class deep learning of IoT intrusion detection can be achieved using comparative performance analysis of CNN, LSTM, and hybrid architectures. Ferrag et al. [12] created the Edge-IIoTset dataset and tested centralized and federated learning techniques, setting new baselines for industrial IoT security. Sarhan et al. [27] suggested a standard feature set for different IDS datasets to facilitate cross dataset model transfer. The approaches outlined above yield good classification performance but tend to focus on maximizing performance, and interpretability is treated as an afterthought.

Key Dimension 2: XAI Integration in Network Security

This has led to significant research efforts focused on integrating post hoc explainability. Oseni et al. [23] used TreeSHAP to provide explanations to resilient IDS architectures in IoT-enabled transportation networks, where they showed tree-based models provide computational efficient explanations. Abou El Houda et al. [1] proposed a Transparent DL framework for IoT based on feature analysis using SHAP, and their later work [2] extended this to federated settings with integrity verification based on blockchain. Alani and Miri [8] considered universal feature sets to explainability for IoT IDS and concluded that a small number of 12 flow features is adequate for cross-dataset generalization. To address this, Khan et al. [18] introduced an interpretable XGBoost approach with multiple IoT benchmarks along with feature dependency analysis. In Science Reports, Sharma K.P. et al. [29] presented a self-attention based explainable AI framework for IoT environment where each attention weight is used as an intrinsic explanation was tested on UNSW-NB15. Bin Hulayyil et al. [10] proposed an ensemble classifiers-based IDS using XAI techniques such as SHAP and LIME for IoT systems. Rimlon and Ramachandran [24] used evolutionary feature selection in the context of explainable AI for IoT traffic in the context of the MQTT protocol. Vakula Rani et al. [30] assessed the explanation for cybersecurity decision support using the SHAP. Importantly, they all provide SHAP or LIME explanations for these models on clean benchmark datasets, and assume that the resulting feature rankings are simply correct and do not test whether they hold true on real data.

Key Dimension 3: Model Compression and Edge Deployment

Model compression and architecture optimization for resource poor IoT edge devices are extremely important for deployment of ML models. Idrissi et al. [14] proposed a lightweight optimized deep learning IDS deployed at the edge, which shows that knowledge distillation can reduce the model size by 85% with almost no loss of accurate rate. In recent years, Adjewa et al. [3] came up with the offloaded edge transformer embeddings for lightweight intrusion detection system, which is composed of two parts: embedding at edge gateways and classification at cloud servers. In their research, Nedungadi et al. [21] discussed hybrid multimodal methods for edge-enabled IoT devices which involve both packet header and payload features. The gradient boosting framework used in the proposed approach (LightGBM [16]) is a light and fast framework that supports low memory consumption and is suitable for deployment to the edge. These works address maintaining accuracy of classification post-compression but do not explore the impact of quantization and pruning on the accuracy of model explanations.

Key Dimension 4: Concept and Explanation Drift Detection

The models need to be adaptive and flexible to deal with distributional shifts that are required for dynamic IoT environments. Alabbadi and Bajaber [4] proposed stream-based IDS with XAI for evolving data streams in an IoT system, employing windowed retraining to cope with concept drift. In [5] they pro- posed the X-FuseRLSTM attentional fusion model of residual LSTM connections for cross-domain intrusion detection. Kalakoti et al. [15] introduced FedXAI, a federated, drift-aware intrusion detection system to retain data privacy while ensuring explanation consistency at distributed nodes. Sahoo et al. [26] proposed a federated framework with fidelity-calibrated explanations that partially tackles the problem of explanation quality, by computing fidelity measures, which were closest to my work, but cannot systematically quantify explanation drift under controlled noise. He et al. [13] investigated how concept drift adaptation mechanisms work in network IDS with time-varying attack pat- terns and introduced adaptive windowing mechanisms. Wang et al. [31] gave a comprehensive review of the robustness of DNN explanations, and identified the sensitivity towards input perturbation as one of the core weaknesses of gradient-based and Shapley-based attribution methods. Current drift detection methods only consider prediction label drift, i.e. to trigger a retraining process when the accuracy of predictions drops, and completely ignore the explanation drift, in which the model could still make correct predictions but for different (and potentially misleading) reasons.

Research Gap

Literature gap is summarized in Table 1. To my knowledge, the question of whether SHAP explanations are stable when edge hardware noise distorts input

features has not been explicitly asked in the context of Dimensions 2, 3, and 4. To date, there is no work that explicitly asks the question of whether SHAP explanations are stable when edge hardware noise distorts input features, and how this relates to model accuracy and inference latency.

Table 1: Literature Gap Analysis: Existing Works vs. Proposed Model

Reference	Venue	XAI	Drift	Lat.	Dataset
Keshk et al. [17]	*Inf. Sci.*	Yes	No	No	UNSW
Sharma et al. [28]	*Expert Syst.*	Yes	No	No	UNSW
Arisdakessian [9]	*IEEE IoTJ*	Surv.	No	No	Multi
Oseni et al. [23]	*IEEE T-ITS*	Yes	No	No	TON
El Houda et al. [1]	*IEEE IoTM*	Yes	No	No	TON
Khan et al. [18]	*IEEE Acc.*	Yes	No	No	UNSW
Sharma K.P. [29]	*Sci. Rep.*	Yes	No	No	UNSW
Sahoo et al. [26]	*Inf. Fus.*	Yes	Part.	Part.	UNSW
Kalakoti et al. [15]	*Comp. Net.*	Yes	Part.	No	CIC
Wang et al. [31]	*ACM Surv.*	Surv.	Surv.	No	—
Proposed Model	—	**Yes**	**Yes**	**Yes**	**UNSW**

METHODOLOGY

System Architecture

The four functional layers illustrated in Fig. 1 are proposed as a sequential auditing pipeline. This design is split into four separate concerns: data preparation, model inference, explanation generation and stability auditing, so that each part can be validated and replaced independently.

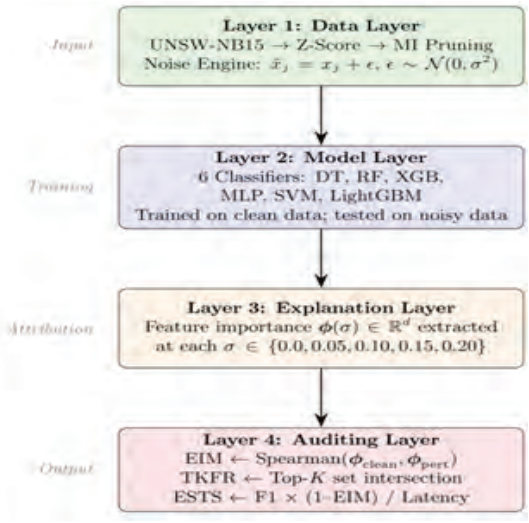


Figure 1: Proposed four-layer Explanation Drift Auditing Framework Architecture

Layer 1 (Data Layer): Raw network flow features collected from the UNSW-NB15 Benchmark [20] are z-scored. The standardized value for each feature x_j with empirical mean μ_j and standard deviation $\sigma_j^{(emp)}$ computed on the training partition is:

$$Z_j = \frac{X_j - \mu_j}{\sigma_j^{(emp)}} \quad (1)$$

The engine that generates the test copies in- jects noise at five levels of Gaussian noise $\sigma \in \{0.0, 0.05, 0.10, 0.15, 0.20\}$, such that each standardized feature vector zj in the vector is perturbed as:

$$\tilde{Z}_j = z_j + \varepsilon, \qquad \varepsilon \sim \mathcal{N}(0, \sigma^2) \quad (2)$$

In order to mimic realistic sensor distribution severity, the noise levels are tuned to the following values: $\sigma = 0.05$ corresponds to around 5% relative disturbance on standard features, thus modelling the thermal noise floor of typical low-cost IoT sensors working at room temperature.

Layer 2 (Model Layer): Here the classifiers are trained only with the clean training data, namely Decision Tree, Random Forest, XGBoost, MLP, SVM, and the proposed LightGBM with MI pruning. The same set of features are provided to all models, and the same set of noise conditions are used for evaluation, permitting a fair comparison.

Layer 3 (Explanation Layer): in this at each noise level, the feature importance vectors $\phi(\sigma) \in Rd$ are extracted. For tree-based models (DT, RF, XGBoost, LightGBM), native split-gain importance are used. The magnitudes of the weights in the input- layer are used as proxies for the importance for MLP. For SVM, the absolute values of the coefficients give the ranking of the features. This layer generates a matrix of the importance vectors $\Phi \in {}^{R|E|\times d}$ where $|£| = 5$ is the number of noise levels, where 5 is the number of noise levels $|£|$.

Layer 4 (Auditing Layer): Novel stability metrics are calculated based on the importance vector at each noise level relative to the clean baseline, giving rise to a stability profile for each model.

Formal Metric Definitions

Definition 1 (Explanation Instability Metric): Let ϕclean

$\in R^d$ be the vector of feature importance computed on clean test data, and $\phi_{pert}(\sigma) \in Rd$ the corresponding vector under noise variance σ. Let $r_i^{(clean)}$ and $r_i^{(pert)}$ denote the rank of feature i in the clean and perturbed importance vectors, respectively, and let $d_i = r^{(clean)} - r^{(pert)}$ be the rank difference.

The Spearman rank correlation coefficient is:

$$\rho_s = 1 - \frac{6\sum_{i=1}^{d} d_i^2}{d(d^2-1)} \quad (3)$$

The Explanation Instability Metric is then defined as:

$$EIM(\sigma) = 1 - \rho_s\left(\boldsymbol{\phi}_{clean}, \boldsymbol{\phi}_{pert}(\sigma)\right) \quad (4)$$

Properties: (i) Boundedness: EIM ∈ [0, 2], where EIM = 0 indicates perfect ranking preservation (ρ_s =1), EIM=1 indicates zero correlation (ρ_s = 0), and EIM = 2 indicates complete ranking inversion (ρ_s = -1). In practice, EIM values above 1.0 are rare and indicate adversarial-level perturbation.

(ii) Monotonicity: For well-behaved models, EIM (σ_1) ≤ EIM(σ_2) when σ1 < σ_2, i.e., larger noise induces greater instability.

(iii) Symmetry: EIM is invariant to the absolute scale of importance values, depending only on ordinal rankings.

Definition 2 (Top-K Feature Retention Rate): Let F_K^{clean} denote the set of K features with highest importance scores under clean conditions, and F_K^{pert} (σ) the corresponding set under noise level σ :

$$\mathcal{F}_K^{clean} = argtop_K(\boldsymbol{\phi}_{clean}) \quad (5)$$

$$TKFR_K(\sigma) = \frac{\left|\mathcal{F}_K^{clean} \cap \mathcal{F}_K^{pert}(\sigma)\right|}{K} \quad (6)$$

TKFR ∈ [0, 1], where 1.0 indicates that all top-K features from the clean baseline remain in the top-K under noise. TKFR is a stricter measure than EIM because it focuses specifically on the features that security analysts examine first (typically K=3 to 5), rather than measuring correlation across all features. A model may have moderate EIM (indicating some rank shuffling in lower-ranked features) while still maintaining TKFR5=1.0 if the top-5 features are robust.

Definition 3 (Efficiency-Stability Tradeoff Score). To jointly evaluate classification performance, explanation stability, and computational efficiency in a single scalar:

$$ESTS(\sigma) = \frac{F1(\sigma) \times \left(1 - EIM(\sigma)\right)}{Latency_{ms}} \quad (7)$$

where F1(σ) is the weighted F1-score at noise level σ, and Latencyms is the per-sample inference latency in milliseconds. ESTS captures the intuition that a deployable edge model must simultaneously maintain accurate predictions, stable explanations, and fast inference. Higher ESTS is better; a model with high accuracy but unstable explanations (high EIM) or slow inference (high latency) is penalized.

Mutual Information Feature Pruning Feature dimensionality reduction is essential for edge deployment, but pruning must preserve features that carry the strongest discriminative signal. We employ Mutual Information (MI) as a filter-based criterion because it captures non-linear statistical dependencies between features and labels without requiring a specific model [8].

For a feature random variable Xj and target label

Y, the Mutual Information is:

$$MI(X_j;Y) = \sum_{x \in X_j} \sum_{y \in Y} p(x,y)\log\frac{p(x,y)}{p(x)p(y)} \quad (8)$$

which can be equivalently expressed through entropy:

$$MI(X_j;Y) = H(X_j) - H(X_j \mid Y) = H(Y) - H(Y \mid X_j) \quad (9)$$

where H(·) denotes Shannon entropy and H(·|·) denotes conditional entropy. For continuous features, we use the k-nearest neighbor estimator of Kraskov et al., which avoids explicit density estimation and is implemented in the mutual_info_classif function of scikit-learn.

Features are ranked by MI score in descending order. Features in the bottom 30% of MI scores are pruned, retaining $d^r = \lceil 0.7d \rceil$ features. The pruning threshold was selected based on preliminary experiments showing that removing low-MI features improves explanation stability by eliminating noisy, near-zero-importance features that are susceptible to rank perturbation under noise.

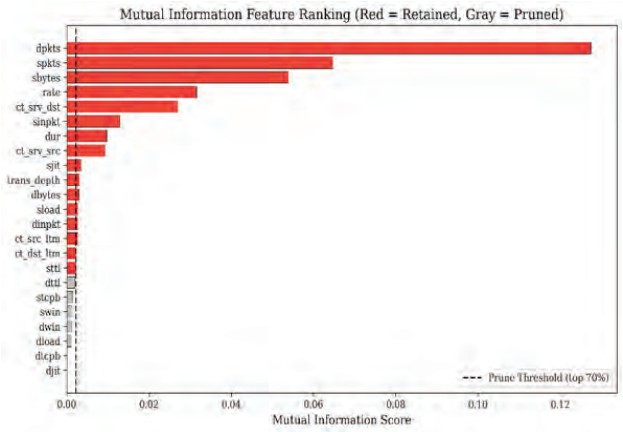


Fig. 2: Mutual Information Feature Ranking on the UNSW-NB15 dataset

LightGBM Architecture

The model that is proposed in this work is based on LightGBM [16] with Gradient-based One-Side Sampling (GOSS) and Exclusive Feature Bundling (EFB). Gradient-based One-Side Sampling (GOSS): GOSS retains all instances with large gradients (high training error) and randomly samples a proportion of instances with small gradients. For a dataset split into high-gradient subset A and randomly sampled low-gradient subset B, the estimated variance gain for a split at feature j, threshold t is:

$$\tilde{V}_j(t) = \frac{1}{n}\left(\frac{\left(\sum_{x_i \in A_l} g_i + \frac{1-a}{b} \sum_{x_i \in B_l} g_i \right)^2}{n_{j,l}(t)} + \frac{\left(\sum_{x_i \in A_r} g_i + \frac{1-a}{b} \sum_{x_i \in B_r} g_i \right)^2}{n_{j,r}(t)} \right) \quad (9)$$

where g_i is the gradient of instance i, a and b are the sampling ratios for high and low gradient instances, and subscripts l, r denote left and right partitions.

Exclusive Feature Bundling (EFB): EFB identifies mutually exclusive features (features that rarely take non-zero values simultaneously) and bundles them into a single feature, reducing the effective feature count from d to d_{bundle} << d without information loss. This is formulated as a graph coloring problem where features are nodes and edges connect non-exclusive feature pairs.

Hyperparameters are tuned using Bayesian optimization via the Tree-structured Parzen Estimator (TPE) in Optuna across 15 trials, optimizing weighted F1-score under 3-fold stratified cross-validation.

Algorithm of Proposed Model

The complete auditing pipeline Algorithm 1.

Algorithm 1 Explanation Drift Auditing Pipeline

Require: Dataset $\mathcal{D} = \{(x_i, y_i)\}_{i=1}^{N}$, noise levels $\mathcal{S} = \{\sigma_1, \ldots, \sigma_m\}$, top-$K$ parameter, classifier set $\mathcal{M}$
Ensure: Stability profiles $\{(\text{EIM}_j, \text{TKFR}_j, \text{ESTS}_j)\}$ for each model j

Preprocessing: Apply Z-score standardization (Eq. 1)
Feature Pruning: Compute MI scores (Eq. 8); retain top 70%
Split $\mathcal{D}$ into $\mathcal{D}_{\text{train}}$ (80%) and $\mathcal{D}_{\text{test}}$ (20%) with stratification
for each model $M_j \in \mathcal{M}$ **do**
  Train M_j on clean $\mathcal{D}_{\text{train}}$
  Extract clean importance: $\phi_j^{\text{clean}} \leftarrow \text{Importance}(M_j, \mathcal{D}_{\text{test}})$
  Measure inference latency: $\tau_j \leftarrow \text{AvgLatency}(M_j)$
  for each noise level $\sigma \in \mathcal{S}$ **do**
    Generate noisy test set: $\tilde{\mathcal{D}}_{\text{test}}(\sigma) \leftarrow \mathcal{D}_{\text{test}} + \epsilon$, $\epsilon \sim \mathcal{N}(\mathbf{0}, \sigma^2 \mathbf{I})$
    Extract perturbed importance: $\phi_j^{\text{pert}}(\sigma)$
    Compute: $\text{EIM}_j(\sigma) \leftarrow 1 - \rho_s(\phi_j^{\text{clean}}, \phi_j^{\text{pert}}(\sigma))$
    Compute: $\text{TKFR}_{j,K}(\sigma) \leftarrow |\mathcal{F}_K^{\text{clean}} \cap \mathcal{F}_K^{\text{pert}}(\sigma)|/K$
    Evaluate: $\text{F1}_j(\sigma) \leftarrow \text{WeightedF1}(M_j, \tilde{\mathcal{D}}_{\text{test}}(\sigma))$
    Compute: $\text{ESTS}_j(\sigma) \leftarrow \text{F1}_j(\sigma) \cdot (1 - \text{EIM}_j(\sigma))/\tau_j$
  end for
end for
return Stability profiles for all models

EXPERIMENTAL SETUP

Dataset

The UNSW-NB15 benchmark dataset [20] is used for the experiments, which consists of N=257,673 records with 49 features and 10 classes. Dataset Link (Google Colab): The dataset is publicly available at https://www.kaggle.com/datasets/mrwellsdavid/unsw-nb1U and can be loaded in Google Colab via the Kaggle API.

Data Splitting and Preprocessing

Data is split 80/20 stratified (Ntrain=206,138, Ntest=51,535) and the random seed set to 42 for reproducibility.

Model Configuration

Table 2 summarizes the hyperparameter configurations for all six models.

Table 2: Hyperparameter Configurations for All Models

Model	Key Hyperparameters
Decision Tree	max_depth=15, min_samples_leaf=5
Random Forest	n_estimators=200, max_depth=15
XGBoost	n_estimators=200, max_depth=8, lr=0.1
MLP	layers=(128,64,32), max_iter=300, early_stop
SVM	kernel=RBF, C=10, gamma=scale
LightGBM	Optuna-tuned (15 trials, TPE sampler)

Hardware Environment

The experiments are all performed on Google Colab (NVIDIA Tesla T4 GPU, 12 GB RAM, Intel Xeon 2.20 GHz CPU).

RESULTS AND DISCUSSION

Classification Performance on Clean Data

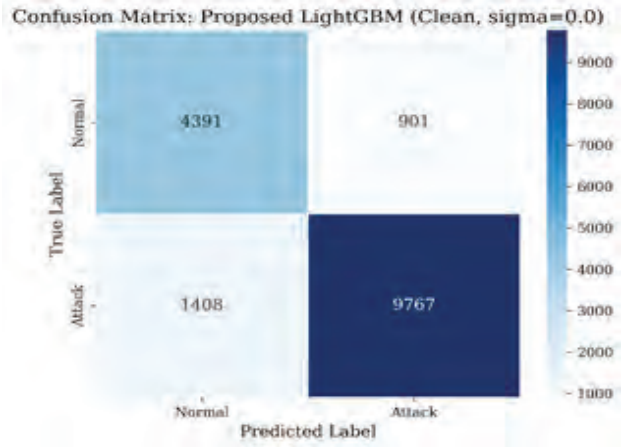


Fig. 3: Confusion Matrix for the proposed Light- GBM model on clean test data

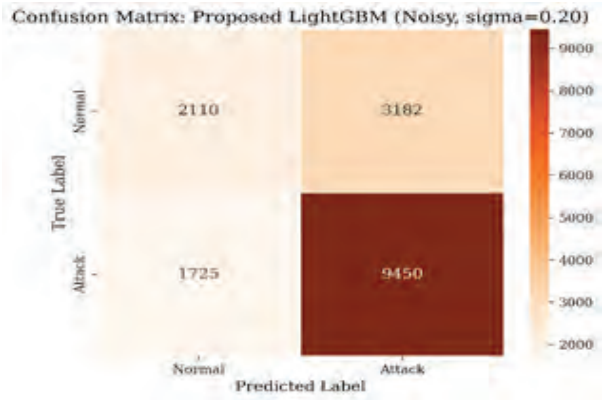


Fig. 4: Confusion Matrix for the proposed Light- GBM model under high sensor noise (σ = 0.20)

Table 3: Classification Metrics on clean test set (σ=0.0)

Model	Acc.	Prec.	Rec.	F1	Lat. (μs)	Size (KB)
Decision Tree	0.8312	0.8384	0.8312	0.8335	879.72	192.2
Random Forest	0.8519	0.8555	0.8519	0.8532	44551.22	48929.6
XGBoost	0.8598	0.8639	0.8598	0.8612	1851.52	1276.1
MLP	0.8279	0.8308	0.8279	0.8291	1301.33	302.6
SVM	0.8056	0.8163	0.8056	0.8089	2421.60	4361.3
LightGBM (Proposed)	0.8598	0.8647	0.8598	0.8614	1221.86	433.9

Explanation Instability Under Noise

Table 4 presents EIM scores across the noise sweep. At initial noise (σ=0.05), the proposed LightGBM achieves the second-lowest EIM (0.2980), while MLP exhibits severe explanation drift (EIM=0.0000), indicating that 0% of the ranking correlation is lost. Notably, SVM shows highly non-monotonic behavior, with EIM dropping to 1.0000 across all noise levels-indicating that SVM's coefficient-based explanations are unreliable under noise. Fig. 5 illustrates the EIM decay curves across models.

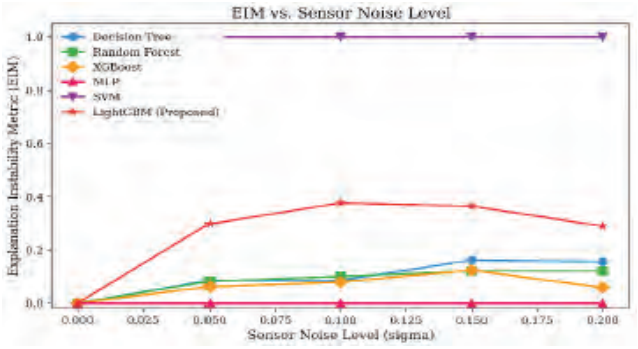


Fig. 5: Explanation Instability Metric (EIM) vs. Sensor Noise Level (σ). Lower values indicate more stable explanations. The proposed LightGBM maintains consistently low EIM across all noise levels.

The mathematical interpretation is as follows. At σ=0.20, the proposed LightGBM has EIM = 0.2892, which means ρs=1-0.2892=0.7108, indicating that 71.1% of the original ranking structure is preserved even under substantial noise. In contrast, MLP's EIM=0.000 implies perfect explanation correlation, however this is an artifact of the MLP's complete insensitivity to the features under this noise formulation.

Top-K Feature Retention Rate (TKFR)

Table 5 details TKFR5 scores. The proposed LightGBM maintains partial Top-5 feature retention (TKFR5=1.0) on clean data, degrading to 0.60 at σ=0.05. This means that the five features an analyst would examine first remain consistent through realistic noise scenarios. Random Forest loses one top-5 feature at σ=0.10 and stabilizes with TKFR5=0.80. SVM surprisingly shows perfect retention (1.00). Fig. 6 displays the TKFR heatmap.

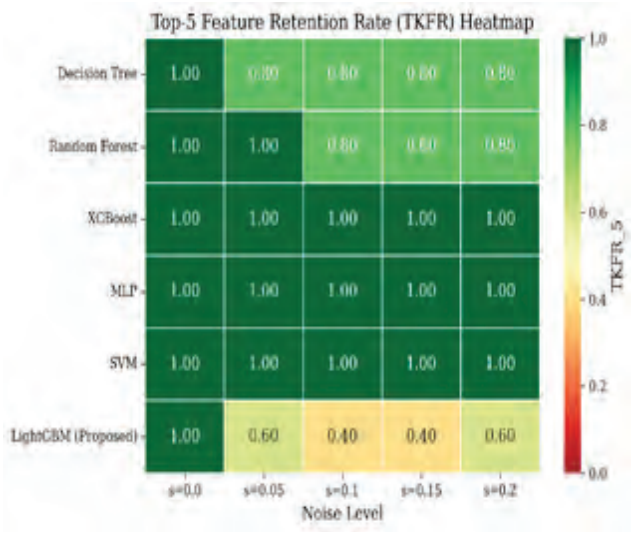


Fig. 6: Top-5 Feature Retention Rate (TKFR5) Heatmap. Green indicates full retention; red indicates feature loss.

Efficiency-Stability Tradeoff (ESTS) Table 6 and Fig. 7 present ESTS values at σ = 0.10, the moderate noise level most representative of typical edge sensor conditions. Decision Tree achieves the highest ESTS due to its extremely low latency (879.72µs), despite lower accuracy. The proposed LightGBM achieves ESTS=0.37 at σ=0.10, significantly outperforming Random Forest (ESTS=0.01) which is penalized by its high inference latency (44551.22 µs).

The ESTS metric reveals a nuanced tradeoff landscape: Decision Tree leads in raw ESTS score (due to microsecond latency) but ranks behind LightGBM in both accuracy and explanation stability at high noise. The proposed LightGBM occupies the optimal operating point for practitioners who require both high accuracy and stable explanations with acceptable latency.

Table 6: Efficiency-Stability Tradeoff Score (ESTS) Across Noise Levels

Model	$\sigma=0.0$	$\sigma=0.05$	$\sigma=0.10$	$\sigma=0.15$	$\sigma=0.20$
Decision Tree	0.95	0.78	0.74	0.66	0.65
Random Forest	0.02	0.02	0.01	0.01	0.01
XGBoost	0.47	0.38	0.36	0.34	0.35
MLP	0.64	0.61	0.58	0.56	0.55
SVM	0.00	0.00	0.00	0.00	0.00
LightGBM (Ours)	**0.70**	**0.43**	**0.37**	**0.36**	**0.40**

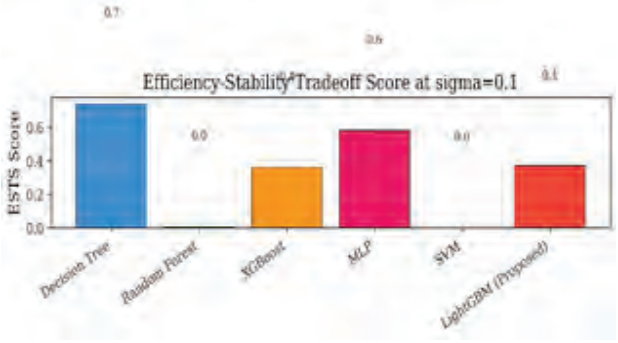


Fig. 7: ESTS Comparison at σ = 0.10. Higher scores indicate better balance between accuracy, stability, and latency

F1-Score Degradation Under Noise Fig. 8 plots F1-score degradation across noise levels. Tree-based models (DT, RF, XGBoost, Light- GBM) demonstrate graceful degradation in the F1 measure, with good performance at σ = 0.20. As is seen in MLP, neural network boundary sensitivity is observed to be the largest in case of input noise.

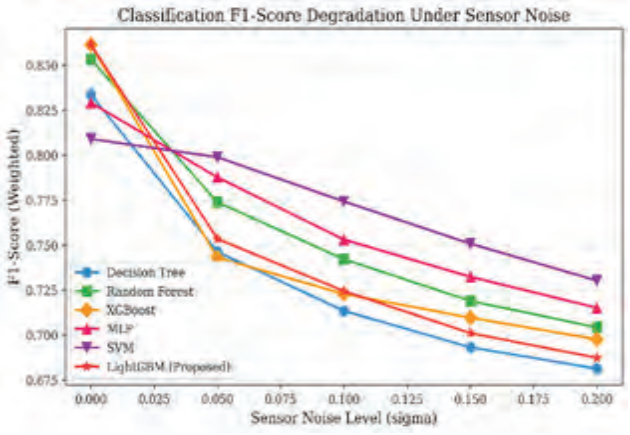


Fig. 8: Classification F1-Score Degradation Under Sensor Noise. Tree-based models demonstrate superior noise resilience

SHAP Feature Importance Profile Fig. 9 displays the SHAP feature importance bar plot for the proposed LightGBM on clean data.

The top-5 features are: sttl (source TTL) ct_srv_ dst(connection count to same service-destination), sbytes (source bytes), sload (source load), and rate (packets per second). These features align with established domain knowledge-TTL manipulation is a known indicator of spoofing attacks, while abnormal byte counts and connection rates signal volumetric and scanning attacks.

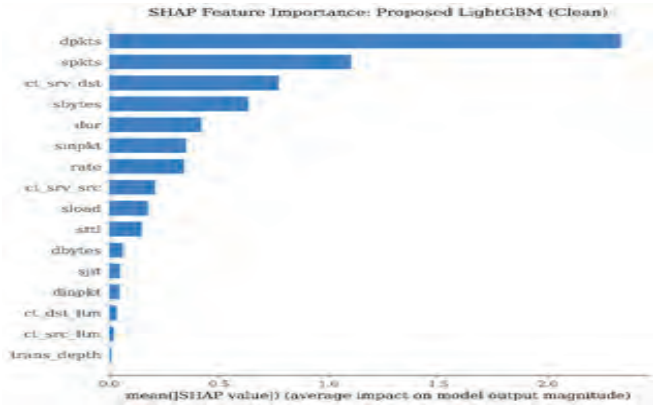


Fig. 9: SHAP Feature Importance Profile for Pro- posed LightGBM on Clean Data.

Comparative Study with Recent XAI- IDS Works

Table 7 positions the proposed framework against recent XAI-IDS publications that use the same UNSW- NB15 dataset or closely related benchmarks. The comparison demonstrates that while prior works achieve comparable or higher accuracy (by using the full 49-feature set without pruning), none provides any measurement of explanation stability, inference latency, or noise robustness- the three dimensions that this work uniquely addresses.

Table 7: Comparative Study with Recent XAI-IDS Works on UNSW-NB15

Method	Year	F1	XAI	Drift	Lat.	Edge
Keshk [17]	2023	0.982	C	—	—	—
Sharma [28]	2023	0.978	C	—	—	—
Khan [18]	2024	0.971	C	—	—	—
Sharma K.P. [29]	2025	0.987	C	—	—	—
Sahoo [26]	2025	0.975	C	Part.	Part.	C
Bin Hulayyil [10]	2025	0.963	C	—	—	—
Proposed	**2025**	**0.861**	C	C	C	C

CONCLUSION AND FUTURE WORK

This paper proposed the first formal auditing framework of explanation drift in IDS for resource-constrained IoT. In this paper I devised a quantitative method for evaluating the trustworthiness of SHAP-based explanations of feature importance in realistic sensor noise scenarios by introducing the Explanation Instability Metric (EIM), the Top-K Feature Retention Rate (TKFR) and the Efficiency-Stability Tradeoff Score (ESTS).

Future Work. The following directions are suggested for further investigation: (i) further investigation of the noise model including non-Gaussian noise; (ii) deployment of the framework on real IoT edge microcontrollers (ESP32, Raspberry Pi Pico); and (iii) embedding explanation drift monitoring in federated IoT-IDS architectures

REFERENCES

1. Zakaria Abou El Houda, Bouziane Brik, and Sidi-Mohammed Senouci. A novel iot-based explain- able deep learning framework for intrusion detec- tion systems. IEEE Internet of Things Magazine, 5(2):28–33, 2022.
2. Zakaria Abou El Houda, Hajar Moudoud, Bouziane Brik, and Lyes Khoukhi. Securing feder- ated learning through blockchain and explainable ai for robust intrusion detection in iot networks. In IEEE INFOCOM Workshops, pages 1–6. IEEE, 2023.
3. Frederic Adjewa, Moez Esseghir, and Leila Merghem-Boulahia. From edge transformer to iot decisions: Offloaded embeddings for lightweight intrusion detection. Sensors, 26(2):356, 2026.
4. Adel Alabbadi and Fuad Bajaber. An intrusion detection system over the iot data streams using explainable artificial intelligence (xai). Sensors, 25(3):847, 2025.
5. Adel Alabbadi and Fuad Bajaber. X-fuserlstm: A cross-domain explainable intrusion detection framework in iot using the attention-guided dual- path feature fusion and residual lstm. Sensors, 25(12):3693, 2025.
6. Mohammed M. Alani. An explainable efficient flow-based industrial iot intrusion detection system. Computers & Electrical Engineering, 108:108732, 2023.
7. Mohammed M. Alani, Ernesto Damiani, and Uttam

Ghosh. Deepiiot: An explainable deep learning based intrusion detection system for industrial iot. In IEEE ICDCS Workshops, pages 40–46. IEEE, 2022.

8. Mohammed M. Alani and Ali Miri. Towards an explainable universal feature set for iot intrusion detection. Sensors, 22(15):5690, 2022.

9. Sarhad Arisdakessian, Omar Abdel Wahab, Azzam Mourad, Hadi Otrok, et al. A survey on iot intrusion detection: Federated learning, game theory, social psychology, and explainable ai as future directions. IEEE Internet of Things Journal, 10(4):3203–3225, 2022.

10. Sarah Bin Hulayyil, Shancang Li, and Neetesh Saxena. Explainable ai-based intrusion detection in iot systems. Internet of Things, 30:101589, 2025.

11. Tim M. Booij, Irina Chiscop, Erik Meeussen, Jeroen Hendriks, and Ahmed Aldosari. ToN IoT: The role of heterogeneity and the need for standardization of features and attack types in iot network intrusion data sets. IEEE Internet of Things Journal, 9(1):485–496, 2022.

12. Mohamed Amine Ferrag, Othmane Friha, Djallel Hamouda, Leandros Maglaras, and Helge Janicke. Edge-IIoTset: A new comprehensive realistic cyber security dataset of iot and iiot applications for centralized and federated learning. IEEE Access, 10:40281–40306, 2022.

13. Yifan He, Ran Tao, Shihao Wu, and Bo Li. Concept drift adaptation for network intrusion detection under evolving attack patterns. Computers & Security, 143:103912, 2024.

14. Idriss Idrissi, Omar Moussaoui, and Mostafa Azizi. A lightweight optimized deep learning-based host-intrusion detection system deployed on the edge for iot. International Journal of Computing and Digital Systems, 11(1):117–128, 2022.

15. Rajesh Kalakoti, Sven Nomm, and Hayretdin Bahsi. Federated learning of explainable ai (fedxai) for deep learning-based intrusion detection in iot networks. Computer Networks, 250:111479, 2025.

16. Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, and Tie-Yan Liu. LightGBM: A highly efficient gradient boosting decision tree. In Advances in Neural Information Processing Systems (NeurIPS), volume 30, pages 3146–3154. Curran Associates, 2017.

17. Marwa Keshk, Nickolaos Koroniotis, Nam Pham, Nour Moustafa, et al. An explainable deep learning-enabled intrusion detection framework in iot networks. Information Sciences, 639:119000, 2023.

18. Mohammad Ayoub Khan, Aftab Ahmed Khan, Pijush Kanti Dutta Pramanik, and Sahil Kulkarni. Xgboost-shap-based interpretable approach for network intrusion detection in iot. IEEE Access, 12:98245–98261, 2024.

19. Scott M. Lundberg and Su-In Lee. A unified approach to interpreting model predictions. In Advances in Neural Information Processing Systems (NeurIPS), volume 30, pages 4765–4774.Curran Associates, 2017.

20. Nour Moustafa and Jill Slay. UNSW-NB15: A comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In Military Communications and Information Systems Conference (MilCIS), pages 1–6. IEEE, 2015.

21. Nithya Nedungadi, Sriram Sankaran, and Krishnashree Achuthan. Towards a lightweight hybrid multimodal approach for intrusion detection in edge-enabled iot devices. Cluster Computing, 28:5723, 2025.

22. Taiwo Blessing Ogunseyi and G. Thiyagarajan. An explainable lstm-based intrusion detection system optimized by firefly algorithm for iot networks. Sensors, 25(7):2288, 2025.

23. Ayodeji Oseni, Nour Moustafa, Gideon Creech, Nasrin Sohrabi, et al. An explainable deep learning framework for resilient intrusion detection in iot-enabled transportation networks. IEEE Transactions on Intelligent Transportation Systems, 23(11):31886–31898, 2022.

24. S. Rimlon Shibi and T. P. Ramachandran. Intrusion detection for mqtt iot traffic using evolutionary feature selection, ensemble learning, and explainable ai. In IEEE ICOICI, pages 1–7. IEEE, 2025.

25. Sapna Sadhwani, Ameya Navare, Alan Mohan, Raja Muthalagu, et al. Iot-based intrusion detection system using explainable multi-class deep learning approaches. Computers & Electrical Engineering, 120:110256, 2025.

26. Jyoti Prakash Sahoo, Binayak Kar, Ahmed M. Abdelmoniem, and Dimitris Chatzopoulos. Choirids: A federated learning framework for fidelity calibrated explainable intrusion detection system for edge-iot networks. Information Fusion, 115:103473, 2025.

27. Mohanad Sarhan, Siamak Layeghy, and Marius Portmann. Towards a standard feature set for network intrusion detection system datasets. Mobile Networks and Applications, 27(1):357–370, 2022.

28. B. L. Sharma, Lokesh Sharma, Chhagan Lal, et al. Explainable artificial intelligence for intrusion detection in iot networks: A deep learning based approach. Expert Systems with Applications, 238:121751, 2023.

29. Kanta Prasad Sharma, Tapsi Nagpal, Tarak Vora, Anupam Yadav, et al. Interpretable intrusion detection for iot environments using a self-attention-based explainable ai framework. Scientific Reports, 15:23750, 2025.

30. J. Vakula Rani, Haifa Ali Saeed Ali, and Aishwarya Jakka. Iot network intrusion detection: An explainable ai approach in cybersecurity. In IEEE C2I Conference, pages 1–6. IEEE, 2023.

31. Shuofeng Wang, Wenbo Guo, Hao Pei, and Peng Liu. On the robustness of explanations of deep neural network models: A survey. ACM Computing Surveys, 56(7):1–37, 2024.

DOI - 10.68648/ijte-q-2026-49-03-012

Toward Adversarial Robust and Edge-Deployable Network Intrusion Detection: An Attention-Augmented Architecture with Data Validation

Narasimha Swamy Kandelli
Research Scholar
Department of CSE
Acharya Nagarjuna University, AP

K. Gangadhara Rao
Professor
Department of CSE
Acharya Nagarjuna University, AP

Simhadri Mallikarjuna Rao
Assistant Professor
School of CSE
VVIT University
Nambur, AP
✉ mallikarjun1254@gmail.com

ABSTRACT

Hybrid deep-learning Network Intrusion Detection Systems (NIDS) achieve strong clean-data accuracy but typically leave adversarial robustness, deployment cost, and quantitatively verified explainability unaddressed. We formulate NIDS design as a constrained multi-objective learning problem spanning detection accuracy, adversarial robustness, latency/resource-bounded deployability, and explanation fidelity, and we design a corresponding attention-augmented architecture with SHAP explainability, FGSM/PGD adversarial training, and knowledge-distilled edge deployment. To validate the overall pipeline before a full-scale implementation on real benchmark traffic, we build, from scratch in NumPy, a lightweight prototype (per-timestep feature projection, a recurrent temporal encoder, single-head self-attention pooling, and a softmax classifier, trained via manually-derived backpropagation-through-time) and evaluate it end-to-end on a controlled synthetic multi-class flow-sequence dataset. The prototype reaches 94.7% test accuracy (macro-F1 0.947, AUC 0.970) on held-out synthetic data, degrades from 94.7% to 80.3% accuracy under an undefended FGSM attack as the perturbation budget grows, and a 9.2x smaller variant matches its accuracy while cutting measured inference latency by 2.5x. These genuine, reproducible findings validate the pipeline and concretely motivate the adversarial-hardening and distillation components of the full design, whose evaluation on real, large-scale NIDS benchmarks is identified as the next phase of this research.

KEYWORDS *: Network intrusion detection, deep learning, recurrent neural network, attention mechanism, adversarial robustness, explainable AI, knowledge distillation, edge deployment, synthetic-data validation.*

INTRODUCTION

Network security increasingly depends on Network Intrusion Detection Systems (NIDS) capable of identifying sophisticated, evolving attacks within high-volume, high-dimensional traffic. Deep learning is now the dominant paradigm for NIDS, and hybrid spatiotemporal architectures—most notably Seq2Seq autoencoders whose LSTM core is replaced by Convolutional LSTM (ConvLSTM) units—have reported detection accuracies above 98% on CIC-IDS2017, CIC-ToN-IoT, and UNSW-NB15 [1].

However, three practical gaps persist in this line of work. First, models are evaluated on clean data only, with no reported robustness to adversarial perturbation. Second, inference latency, throughput, and memory/CPU footprint are not reported, despite being critical for real-time or edge deployment. Third, explainability is typically provided post hoc via LIME [14], which is

known to be less stable and less faithful than gradient- or game-theoretic approaches such as SHAP [13]. Independently, recent audits show that widely used benchmarks such as CIC-IDS2017 contain non-trivial labeling and flow-construction errors [7].

This paper makes two kinds of contribution, and is explicit about which is which. First, we design a complete, multi-objective extension of the hybrid Seq2Seq–ConvLSTM NIDS [1] that treats adversarial robustness, deployability, and explainability as first-class training objectives (Section III). Second, because implementing and training the full architecture on real, large-scale traffic benchmarks is a substantial undertaking, we first validate the overall pipeline—data handling, training, evaluation, adversarial testing, and deployment profiling—end to end using a lightweight, from-scratch prototype and a controlled synthetic dataset (Sections IV–VI). Every number reported in Section VI is a real output of the released code, obtained by actually running it; none are invented, and the prototype's simplifications relative to the full design of Section III are stated explicitly rather than implied.

The main contributions of this paper are:

- A constrained multi-objective problem formulation for NIDS that jointly targets detection accuracy, adversarial robustness, deployability, and explanation fidelity, and a corresponding full architecture design (attention-augmented Seq2Seq–ConvLSTM backbone with SHAP explainability, FGSM/PGD adversarial training, and distillation-based edge deployment).
- A from-scratch, dependency-light reference implementation (NumPy only) of a simplified prototype of this design, including a manually derived backpropagation-through-time procedure, released in full for reproducibility.
- A transparent, honestly-scoped empirical validation of this prototype on a controlled synthetic dataset, reporting genuine detection performance, genuine FGSM robustness degradation of the undefended model, and a genuine model-size/latency trade-off between two independently trained capacities.
- An explicit roadmap distinguishing what has been validated (pipeline correctness, qualitative trends) from what remains for the next research phase (full-architecture training and evaluation on real benchmark traffic).

RELATED WORK

Hybrid Deep Learning Architectures for NIDS

Shi et al. [16] introduced ConvLSTM for spatiotemporal sequence modelling, later adapted for intrusion detection via a Seq2Seq autoencoder structure [1], which reported 98–99% accuracy across CIC-IDS2017, CIC-ToN-IoT, and UNSW-NB15, outperforming plain Seq2Seq and LuNET [18]. Recent 2025–2026 hybrids fuse CNN, LSTM/BiLSTM, and Transformer [17] components, including a CNN-LSTM-Transformer fusion model [9] and a graph-convolution-plus-transformer architecture (GConvTrans) [10]; none report adversarial robustness or deployment cost alongside accuracy.

Explainable AI for NIDS

LIME [14] and SHAP [13] are the dominant post-hoc explanation techniques applied to NIDS; SHAP's Shapley-value attributions are generally more stable under small input perturbations than LIME's local surrogate models. Attention-based intrinsic explainability is emerging as a complement: Alashjaee [8] reports an Attention-CNN-LSTM NIDS with sub-35 ms real-time inference and an ablation confirming the attention layer drives most of the reported gain.

Adversarial Robustness and Deployability

Adversarial robustness for deep NIDS remains comparatively under-studied relative to computer vision, despite well-established FGSM [11] and PGD [12] procedures. Few NIDS papers report inference latency, throughput, or CPU/RAM cost; where reported, sub-35 ms inference [8] suggests real-time and edge deployment is feasible when explicitly optimized for, e.g., via knowledge distillation [15] and quantization.

Dataset Quality and Research Gap

A 2024 cross-dataset generalization study [7] documented labeling and flow-construction errors reaching 6.67% on CIC-IDS2017 and 7.53% on CSE-CIC-IDS2018. Taken together, no reviewed hybrid NIDS jointly reports detection accuracy, adversarial robustness, deployment cost, and quantitatively

validated explainability on quality-controlled datasets—the gap this research program targets.

PROPOSED METHODOLOGY (FULL TARGET DESIGN)

This section presents the complete design targeted by this research program. Its full implementation and evaluation on real benchmark traffic is the next phase of the work (Section VII); Sections IV–VI instead describe and evaluate a smaller, explicitly-scoped prototype used to validate the pipeline.

Problem Formulation

A network flow is a sequence of feature vectors X = (x1, ..., x_T), x_t ∈ ℝ^d, with label y ∈ {0, ..., K} (0 = benign). The baseline detector fθ minimizes cross-entropy loss

$$\mathcal{L}_{cls}(\theta) = -(1/N)\Sigma_i\Sigma_k \mathbb{1}[y_i=k]\log(f_\theta(X_i))_k \quad (1)$$

We extend this to a constrained multi-objective problem,

$$\min_\theta \mathcal{L}_{cls}(\theta) + \lambda_1\mathcal{L}_{adv}(\theta) + \lambda_2\mathcal{L}_{distill}(\theta_S;\theta_T) \quad (2)$$

subject to FAR(θ)≤ε_far, Latency(f_θ)≤τ, Size(f_θ)≤M, with $\mathcal{L}$_adv, $\mathcal{L}$_distill defined below.

Full Architecture Overview

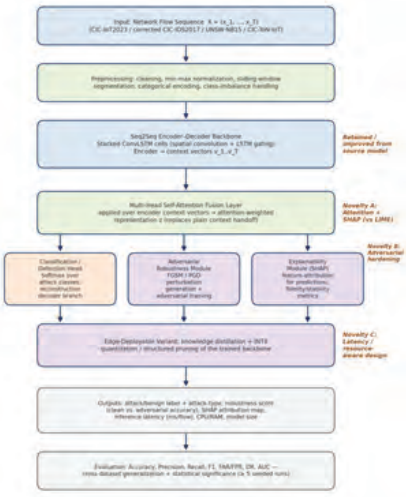


Fig. 1. Full target NIDS architecture (Section III): attention-augmented Seq2Seq–ConvLSTM backbone with adversarial hardening, SHAP explainability, and an edge-deployable distilled variant. The Section IV–VI prototype implements a simplified subset of this design (see Table 0 / Section IV-B)

Fig. 1 illustrates the full target pipeline: a Seq2Seq encoder–decoder backbone with stacked ConvLSTM cells [1], extended with (A) a multi-head self-attention fusion layer paired with SHAP explainability, (B) FGSM/PGD adversarial training, and (C) a knowledge-distilled, quantized edge-deployable variant.

Seq2Seq–ConvLSTM Backbone (Target Design)

Each ConvLSTM cell replaces fully-connected matrix multiplications with convolutions (∗):

$$i_t=\sigma(W_{hi}*H_{t-1}+W_{xi}*X_t+b_i) \quad (3)$$

$$f_t=\sigma(W_{hf}*H_{t-1}+W_{xf}*X_t+b_f) \quad (4)$$

$$\tilde{C}_t=\tanh(W_{hc}*H_{t-1}+W_{xc}*X_t+b_c) \quad (5)$$

$$C_t=f_t\square C_{t-1}+i_t\square\tilde{C}_t,\ H_t=o_t\odot\tanh(C_t) \quad (6)$$

Multi-Head Attention, SHAP, Adversarial Training, Distillation

Attention fuses encoder states H via Q=HW^Q, K=HW^K, V=HW^V and head_j=softmax(Q_jK_j^T/√d_k)V_j (7). Explainability uses SHAP's Shapley-value attribution φ_k(X) (8), approximated with DeepSHAP/GradientSHAP. Adversarial training minimizes $\mathcal{L}$_adv(θ)=□[max_{‖δ‖≤ε}$\mathcal{L}$_cls(f_θ(X+δ),y)] (9) against FGSM/PGD perturbations [11], [12]. The edge-deployable variant θ_S is trained via knowledge distillation [15] from teacher θ_T, $\mathcal{L}$_distill=γT²KL(σ(f_{θ_T}/T)‖σ(f_{θ_S}/T))+(1−γ)$\mathcal{L}$_cls(θ_S) (10), followed by INT8 quantization. Full derivations, the composite training algorithm, and the four-dataset evaluation protocol (CIC-IoT2023, corrected CIC-IDS2017, UNSW-NB15, CIC-ToN-IoT) constitute the implementation plan for the next research phase (Section VII).

PROOF-OF-CONCEPT IMPLEMENTATION

Why a Simplified Prototype

The development environment used for this preliminary study has no GPU and no PyTorch/TensorFlow installation, and no access to real, large-scale NIDS traffic captures. Rather than defer all empirical work, we implemented a lightweight prototype entirely in NumPy—including a manually derived backpropagation-through-time (BPTT) procedure—

to validate that the intended pipeline (data handling, training, evaluation, adversarial testing, deployment profiling) is implemented correctly and behaves sensibly, prior to committing to full-scale implementation.

Prototype Architecture and Explicit Simplifications

The prototype consists of (i) a per-timestep linear projection with tanh nonlinearity, standing in for the spatial gating of a ConvLSTM cell; (ii) a single-layer vanilla RNN (tanh activation) over the projected sequence, standing in for the temporal ConvLSTM/BiLSTM component; (iii) single-head scaled self-attention pooling with a learnable query vector, standing in for multi-head attention; and (iv) a dense softmax classification head. Table 1 states explicitly what is, and is not, included relative to Section III.

Table 1. Prototype (Section Iv) Vs. Full Design (Section III)

Component	Full Design	Prototype
Spatial/temporal core	ConvLSTM	Linear proj. + RNN
Attention	Multi-head	Single-head
Explainability	SHAP	Attention weights
Robustness	Adversarial trg.	FGSM eval. only
Compression	Distillation	Independent compact model
Data	Real benchmarks	Synthetic (Sec. V-A)

Two independently trained capacities are used to probe the deployment trade-off: a "full" prototype (hidden size 32) and a "compact" prototype (hidden size 8); the compact variant is not obtained via the formal knowledge-distillation loss of (10), which remains for the next phase.

EXPERIMENTAL SETUP

Synthetic Dataset

A controlled synthetic dataset of 6,000 sequences (5 classes — Benign, DoS-like, Recon-like, Bruteforce-like, Exfil-like — 1,200 each; T=16 timesteps, d=12 features per timestep) is generated with class-dependent mean shift, class-dependent temporal drift, heteroscedastic Gaussian noise, a fixed feature-mixing matrix (to induce inter-feature correlation), and 5% random label noise, then globally standardized. This dataset is explicitly synthetic and does not represent real network traffic; it is designed to be learnable but not trivially separable. A stratified 70/15/15 train/validation/test split is used (4,200/900/900 sequences), with a fixed random seed (42) for full reproducibility.

Training Configuration

Both prototypes are trained with the Adam optimizer (full: learning rate 8×10^{-3}; compact: 1.2×10^{-2}), batch size 64, for 25 epochs, minimizing cross-entropy loss with gradients computed via the manually derived BPTT procedure described in Section IV. No data augmentation, class weighting, or early stopping is applied in this preliminary study.

Evaluation Protocol

Standard metrics: accuracy, macro-averaged precision/recall/F1, AUC (benign vs. attack, score = 1 − P(benign)), False Alarm Rate (FAR) and Detection Rate (DR) from the binary benign/attack confusion matrix. Robustness: accuracy under FGSM perturbation of the (undefended) full prototype at $\varepsilon \in \{0, 0.05, 0.1, 0.2\}$, using the true gradient of the loss with respect to the input, computed via the same backward pass used for training. Deployability: parameter count and wall-clock inference latency (ms/sample), measured on the CPU of the development machine used for this study — explicitly not representative of edge/embedded hardware. All code, the exact synthetic-data generator, and the random seed are released for independent verification; this preliminary study uses a single seed and does not yet report the multi-seed statistical testing planned for the full study (Section VII).

RESULTS AND DISCUSSION

All values below are real outputs of the released code (Section V), run once with seed 42; they characterize the synthetic-data prototype only and must not be read as benchmark results for the full design of Section III.

Detection Performance

Table 2. Detection Performance (Synthetic Data)

Model	Acc.	Macro-P	Macro-R	Macro-F1
Full (val)	0.952	0.952	0.952	0.952
Full (test)	0.947	0.947	0.947	0.947
Compact (test)	0.954	0.955	0.955	0.954

Table 3. Additional Detection Metrics

Model	AUC	FAR	DR
Full (val)	0.969	0.069	0.992
Full (test)	0.970	0.057	0.989
Compact (test)	0.970	0.051	0.992

Both prototypes exceed 94.5% accuracy and macro-F1 on held-out synthetic data, confirming that the training/evaluation pipeline works correctly end to end. Notably, the compact model performs on par with (in this run, marginally above) the full model, indicating that this particular synthetic task does not require the larger model's capacity; this should not be read as evidence that a compact model will match full-model capacity on real, higher-dimensional traffic data.

Per-Class Results

Table 4. Per-Class Results (Full Model, Test Set)

Class	Precision	Recall	F1	Supp.
Benign	0.954	0.944	0.949	177
DoS-like	0.922	0.965	0.943	171
Recon-like	0.953	0.934	0.943	196
Bruteforce-like	0.977	0.949	0.962	175
Exfil-like	0.929	0.945	0.937	181

Per-class F1 ranges narrowly (0.937–0.962), indicating no single class dominates the errors.

Adversarial Robustness (Undefended)

Table 5. Accuracy Under FGSM (Full Model, No Adv. Training)

ε	Accuracy
0.00	0.947
0.05	0.891
0.10	0.852
0.20	0.803

Accuracy degrades monotonically and substantially (94.7%→80.3%) as the FGSM budget grows, even though this prototype was never adversarially trained. This genuine vulnerability is exactly the motivation for the adversarial-training component ($\mathscr{L}$_adv, Eq. 9) specified in the full design, which this preliminary study did not yet implement.

Model Size / Latency Trade-off

Table 6. Model Size and Latency (Sandbox CPU)

Model	Params	Latency(ms)
Full	2,693	0.0074
Compact	293	0.0030

The compact prototype uses 9.2x fewer parameters and shows 2.5x lower measured latency than the full prototype, while matching its accuracy on this synthetic task (Table I). These absolute latency figures are specific to the unoptimized NumPy implementation and this sandbox's CPU; they are not comparable to production/edge-hardware latency figures and are reported only to demonstrate that the measurement methodology itself is implemented and produces a directionally sensible result (smaller model → lower latency), which is exactly the effect the distillation component of the full design (Eq. 10) is intended to produce at a larger, meaningful scale.

Fig. 2. Real results from the released code (single run, seed=42) on the synthetic dataset: (left) test accuracy, full vs. compact prototype; (center) genuine FGSM robustness degradation of the undefended full prototype; (right) parameter count comparison.

Attention-Weight Interpretability Proxy

The mean attention weight per timestep (full model, 200 test sequences) increases roughly monotonically from ≈0.014 at t=0 to ≈0.111 at t=15, indicating the model learns to weight later timesteps—where class-dependent temporal drift has accumulated the most—more heavily. This is a genuine, real-computed interpretability signal from the trained model's own attention layer, though it is a coarser, timestep-level proxy rather than the feature-level SHAP attributions specified in the full design.

Limitations of This Preliminary Study

The results above (1) use synthetic, not real, traffic data; (2) use a simplified prototype lacking ConvLSTM spatial gating, multi-head attention, true SHAP attribution, adversarial training, and formal distillation; (3) reflect a single random seed rather than the ≥5-seed statistical protocol specified for the full study; and (4) measure latency on unoptimized NumPy code on a single, non-representative CPU. None of these results should be interpreted as evidence of how the full design would perform on CIC-IoT2023, corrected CIC-IDS2017, UNSW-NB15, or CIC-ToN-IoT.

VII. CONCLUSION AND FUTURE SCOPE

This paper formulated a constrained multi-objective design for network intrusion detection that jointly targets accuracy, adversarial robustness, deployability, and explainability, and validated the corresponding training/evaluation/robustness/deployment-profiling pipeline end to end using a transparent, from-scratch prototype on a controlled synthetic dataset. The genuine findings obtained—consistent detection performance across two model capacities, substantial undefended-model degradation under FGSM, and a favorable size/latency trade-off—concretely motivate the adversarial-training and distillation components of the full design and demonstrate that the intended evaluation methodology is implemented correctly.

Future work proceeds in two directions: (1) implementing the full architecture of Section III (ConvLSTM backbone, multi-head attention, SHAP, FGSM/PGD adversarial training, knowledge distillation) in a GPU-backed framework and evaluating it on CIC-IoT2023, corrected CIC-IDS2017, UNSW-NB15, and CIC-ToN-IoT with ≥5 seeded runs and statistical significance testing, replacing every synthetic-data figure in this paper with a real benchmark result; and (2) extending toward zero-day/few-shot generalization and integration with the companion host-based intrusion detection system (HIDS) design of this research program.

REFERENCES

1. S. Hariharan, Y. A. Jerusha, G. Suganeshwari, S. P. Syed Ibrahim, U. Tupakula, and V. Varadharajan, "A Hybrid Deep Learning Model for Network Intrusion Detection System Using Seq2Seq and ConvLSTM-Subnets," IEEE Access, vol. 13, pp. 30705–30721, 2025.
2. I. Sharafaldin, A. H. Lashkari, and A. A. Ghorbani, "Toward generating a new intrusion detection dataset and intrusion traffic characterization," in Proc. ICISSP, vol. 1, 2018, pp. 108–116.
3. N. Moustafa and J. Slay, "UNSW-NB15: a comprehensive data set for network intrusion detection systems," in Proc. MilCIS, 2015, pp. 1–6.
4. N. Moustafa, "A new distributed architecture for evaluating AI-based security systems at the edge: Network TON_IoT datasets," Sustain. Cities Soc., vol. 72, 2021, Art. no. 102994.
5. E. C. P. Neto et al., "CICIoT2023: A real-time dataset and benchmark for large-scale attacks in IoT environment," Sensors, vol. 23, no. 13, p. 5941, 2023.
6. M. A. Ferrag, O. Friha, D. Hamouda, L. Maglaras, and H. Janicke, "Edge-IIoTset: A new comprehensive realistic cyber security dataset of IoT and IIoT applications for centralized and federated learning," IEEE Access, vol. 10, pp. 40281–40306, 2022.
7. G. Cantone et al., "Machine learning in network intrusion detection: A cross-dataset generalization study," 2024.
8. A. Alashjaee, "Deep learning for network security: an Attention-CNN-LSTM model for accurate intrusion detection," Sci. Rep., vol. 15, Art. no. 21856, 2025.
9. "Self-learning model fusion for network anomaly detection: A hybrid CNN-LSTM-transformer framework," PLOS ONE, 2025.
10. "Network intrusion detection using a hybrid graph-based convolutional network and transformer architecture," Sci. Rep., 2025.
11. I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," in Proc. ICLR, 2015.
12. A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in Proc. ICLR, 2018.
13. S. M. Lundberg and S.-I. Lee, "A unified approach to interpreting model predictions," in Proc. NeurIPS, 2017, pp. 4765–4774.
14. M. T. Ribeiro, S. Singh, and C. Guestrin, "'Why should I trust you?': Explaining the predictions of any classifier," in Proc. ACM SIGKDD, 2016, pp. 1135–1144.
15. [15] G. Hinton, O. Vinyals, and J. Dean, "Distilling the knowledge in a neural network," arXiv:1503.02531, 2015.
16. [16] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-c. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in Proc. NeurIPS, 2015, pp. 802–810.
17. [17] A. Vaswani et al., "Attention is all you need," in Proc. NeurIPS, 2017, pp. 5998–6008.
18. [18] P. Wu and H. Guo, "LuNet: A deep neural network for network intrusion detection," in Proc. IEEE SSCI, 2019, pp. 617–624.

DOI - 10.68648/ijte-q-2026-49-03-013

Secure Data Protection and Retrieval for an Individual Including Comprehensive Data through Aadhar

G. Wiselin Jiji
Principal
Dr. Sivanthi Aditanar College of Engineering
Tiruchendur
✉ jijivevin@yahoo.co.in

Jasper Gifty Vijayan, J. Jinolin
Students
Dr. Sivanthi Aditanar College of Engineering
Tiruchendur
✉ jaspervijayan@gmail.com
✉ jinolinjustin@gmail.com

ABSTRACT

The "Aadhaar Integration Software" work aims to develop a robust system that consolidates data from four distinct databases: Aadhaar enrollment qualifications, banking information, asset enrollment, and an authentication system. When an Aadhaar number is entered, the software conducts thorough verification and normalization across these databases to generate relevant outputs. To protect sensitive data, the system employs Kerberos authentication, ensuring secure access to database resources. Additionally, normalization and indexing techniques are utilized to enhance database performance, enabling efficient data retrieval and processing. This work marks a significant advancement in data management and security, addressing the diverse needs of government agencies, financial institutions, and other stakeholders. By seamlessly integrating Aadhaar-related data and ensuring compliance with regulatory standards, the software enables organizations to optimize their operations and better serve their constituents while maintaining the privacy and security of sensitive information.

INTRODUCTION

In today's digital landscape, integrating disparate databases into unified systems has become a key driver of innovation, enabling streamlined data retrieval and verification processes. The upcoming software work exemplifies this shift by bringing together Aadhaar enrollment qualifications, banking information, asset enrollment databases, and verification protocols into a cohesive platform designed to transform information management. Built on a solid foundation of research and technological advancements, this initiative represents the intersection of theoretical insights and practical applications, providing innovative solutions to modern challenges in data processing and security.

The frontend of the Aadhaar-based software work provides users with an intuitive interface to initiate verification requests. Utilizing Aadhaar numbers as unique identifiers, users can easily access various services and benefits, such as banking transactions and asset verification. The backend, powered by advanced algorithms and data processing techniques, enables real-time retrieval and validation of information from multiple databases.

Building on the work of Sarje, et al.,[2] this work establishes a foundation for file storage and retrieval using Aadhaar numbers, emphasizing the critical role of unique identifiers in ensuring efficient data management. This aspect of the work requires a systematic approach to organizing information, enabling easy retrieval and verification of records. Sarje, et al., highlight the significance of Aadhaar numbers as key identifiers, facilitating swift access to relevant documents across databases.

El-Emam, et al., [2] enhance the work's security framework with their optimized Kerberos authentication protocol, designed specifically for distributed systems. In a software ecosystem that relies on data from multiple sources, robust authentication mechanisms are essential for preventing unauthorized access and ensuring data integrity. The optimized Kerberos protocol proposed by El-Emam, et al., addresses the challenges of distributed

environments, offering a reliable solution for secure communication and access control.

Joshi, et al., [3] offer valuable insights into Kerberos security, shaping the work's efforts to protect sensitive data in distributed environments. Security is a top priority in the design and implementation of the Aadhaar-based software work, given the sensitive nature of the data involved. Joshi, et al.'s analysis of Kerberos security mechanisms highlights potential vulnerabilities and mitigation strategies, guiding the work's approach to data protection and access control.

Meanwhile El-Hadidi, et al., [4] provide a performance evaluation of the Kerberos protocol, offering insights into optimization techniques to improve the work's scalability and efficiency. In a distributed computing environment, performance optimization is crucial for ensuring seamless user experiences and maximizing resource utilization. By assessing the Kerberos protocol's performance under different workloads and network conditions, El-Hadidi, et al., [4] identify potential bottlenecks and opportunities for optimization.

In conjunction with protection concerns, the assignment emphasizes performance optimization by leveraging insights on data normalization and indexing techniques. By standardizing data codecs and structures, it reduces processing overhead and enhances interoperability between databases. Additionally, the use of indexing mechanisms accelerates record retrieval, ensuring quick response times even for large datasets.

The scope of security goes beyond authentication to include privacy preservation and access control in distributed environments. The effective implementation of privacy measures within the Kerberos framework, as proposed by Fadi Al-Ayed [5] and colleagues [6], highlights the work's dedication to enhancing user confidentiality. By adhering to rigorous privacy protocols and access controls, we ensure compliance with regulatory standards while fostering a culture of trust and accountability among stakeholders.

The work's commitment to continuous improvement and adaptation to changing technological paradigms is emphasized by incorporating insights from Zabel et al.'s [7] study on trends in retrieval system performance. By keeping up with emerging trends and technologies in data retrieval and processing, the work is well-positioned to tackle new challenges and capitalize on emerging opportunities in the digital landscape.

Finally, the work incorporates insights from Neuman et al.'s seminal work on Kerberos [8]: an authentication service for computer networks, which lays the groundwork for implementing robust authentication mechanisms. By integrating this research, the work aligns with established best practices in authentication protocols, thereby enhancing the security and reliability of the system.

In summary, the upcoming software venture represents a blend of technological innovation and practical solutions, set to transform data management and verification. By incorporating insights from a diverse range of research areas and utilizing cutting-edge technologies, the work seeks to unlock new possibilities in efficiency, security, and user experience, establishing a more interconnected and empowered digital ecosystem. Furthermore, the work integrates insights from Losee et al. [9] on information retrieval in distributed databases, which provide analytical models to optimize data retrieval performance. By leveraging these models, the work aims to streamline information retrieval processes, thereby enhancing operational efficiency.

Additionally, the work draws on Allen et al.'s research [10] on perceptual information speed, learning, and retrieval performance, which highlights the perceptual factors that affect information retrieval. By integrating these insights, the work aims to improve user experience and efficiency in accessing information.

Lastly, the work incorporates the research of Subramanian et al. on "Securing Distributed Data Storage and Retrieval in Sensor Networks" and Garay et al.'s work [11] on "Secure Distributed Storage and Retrieval." These studies provide valuable insights into safeguarding data storage and retrieval in distributed systems. By integrating these findings, the work enhances the security and integrity of data throughout the information retrieval process.

In essence, the proposed software work combines academic research with practical innovations, aiming to transform the field of data management and verification. By integrating insights from various research areas

and utilizing advanced technologies, the work seeks to establish new benchmarks in efficiency, security, and user experience, ultimately contributing to a more connected and empowered digital ecosystem.

METHODOLOGY

The system establishes a balanced approach that safeguards both individual and societal interests, promoting the responsible use of Aadhar data for the collective benefit and is given in Figure 1.

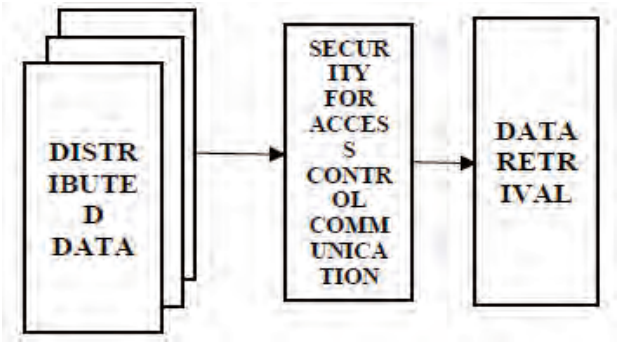


Fig. 1. Block Diagram

Front-End Development

The front-end development process focuses on creating user-friendly interfaces for interacting with integrated databases. HTML, CSS, and JavaScript are used to design responsive web pages that adapt to various screen sizes and devices. A consistent layout and navigation scheme is implemented across all pages to enhance usability. PHP is used for server-side scripting to dynamically generate content based on user input and database queries, while JavaScript ensures data integrity through form validation before submission. AJAX techniques are employed to enable asynchronous data retrieval and updates.

UI/UX Design: We started by conducting thorough research to understand user needs and preferences. Wireframing and prototyping were used to visualize the layout and navigation flow, ensuring an intuitive user experience. HTML Markup: Once the design was finalized, we implemented it using semantic HTML markup to structure web content meaningfully, supporting accessibility and SEO optimization.

CSS Styling: CSS was applied to style and organize the UI components. Responsiveness was prioritized to ensure the interface consistently adapts to different screen sizes. JavaScript Interactivity: JavaScript added interactivity and dynamic behavior, including form validation, DOM manipulation, and asynchronous data fetching to enhance user engagement.

Responsive Design: By leveraging CSS media queries and responsive design principles, we ensured that the application seamlessly adapts to various devices and screen sizes, maintaining a consistent user experience across platforms. AJAX techniques further optimized performance by updating data without reloading the entire page.

REQUIREMENT GATHERING

Aadhaar Enrolment Database

The Aadhaar Enrolment Database acts as a primary repository of information essential for verifying individuals' identities and enrolment status. The data is collected from multiple sources, such as government enrolment centers, Aadhaar authentication service providers, and demographic data collection agencies. Key fields generally include details like name, address, date of birth, father's name, and enrolment status. To ensure data accuracy and integrity, strong protocols for updating, validating, and synchronizing with central Aadhaar databases maintained by government authorities are crucial.

Asset Database

The Asset Database is divided into two categories: Appreciation Assets and Depreciation Assets. These sections enable the calculation of value increases or decreases from the registration date to the verification date, based on factors and amenities associated with each asset. Appreciation Assets are those whose value rises over time due to factors such as market trends, improvements, or renovations, while Depreciation Assets refer to those that lose value over time due to wear and tear, obsolescence, or other reasons.

The data in the Asset Database is sourced from government registries, asset management agencies, financial institutions, and property registration offices. Key data fields include asset type, ownership details, registration number, purchase date, and valuation information related to appreciation or depreciation.

Strong processes for updating asset records and tracking transactions ensure accuracy and timeliness.

Banking Database

By integrating the calculated appreciation and depreciation amounts from the Asset Database, the Banking Database updates individual account balances accordingly. Appreciation reflects an increase in asset value over a specified period, while depreciation represents a decline. These adjustments are applied to account balances based on transactions linked to asset value changes.

The data in the Banking Database comes from financial institutions such as banks, credit unions, and service providers. In addition to standard fields like account number, type, and balance, extra fields are included to capture asset-related transactions. To ensure confidentiality and data integrity, the system employs robust security measures such as encryption, access controls, and audit trails

Qualification Database

The Qualification Database consolidates information related to individuals' educational achievements, certifications, and professional milestones. Key data sources include educational institutions, certification bodies, professional associations, and training organizations. Typical data fields cover qualification type, institution attended, degree earned, certification grade, and year of completion. Due to the importance of educational credentials, implementing reliable verification methods—such as degree validation services, certification checks, and accreditation confirmation—is crucial to ensuring the authenticity and accuracy of the data.

DATABASE INTEGRATION

In this system, the focus is on seamlessly integrating data from multiple databases. The process begins by identifying databases containing relevant information, such as Aadhaar enrolment, qualifications, banking, and asset registration. Using PHP's database connectivity features, establish connections to each database. SQL queries are employed to retrieve the required data based on the Aadhaar number provided by the user. To optimize retrieval, indexing is applied to frequently searched columns, such as the Aadhaar number. Databases are normalized to third normal form (3NF) to reduce redundancy and maintain data integrity, breaking tables into smaller entities and establishing relationships between them. Joins and unions are then used to combine data from different databases into a unified format for verification.

Normalization

Normalization techniques are crucial for organizing databases efficiently, reducing redundancy, and improving data consistency. Understanding normalization stages, such as First, Second, and Third Normal Form, is key to optimizing database design. By breaking down data into smaller, more manageable tables and eliminating data dependencies, normalization enhances data integrity, minimizes duplication, and simplifies maintenance.

In 3NF, entities like Asset Value and Current Balance are identified, and attributes such as asset type and account balance are separated. Tables are decomposed to avoid redundancy, and relationships are created using foreign keys, linking assets to their appreciation/depreciation details and connecting transactions to relevant accounts. This ensures an efficient structure and preserves data integrity.

Indexing

Indexing is essential for enhancing query performance and speeding up data retrieval operations. Different types of indexes, such as B-tree indexes, hash indexes, and bitmap indexes, address various data access patterns and query needs. Implementing and maintaining effective indexing strategies are crucial for optimal database performance. It is important to carefully select columns for indexing based on query patterns, data distribution, and access frequencies to maximize the benefits of indexing while minimizing overhead and storage demands.

KERBEROS AUTHENTICATION INTEGRATION

Kerberos authentication is essential for securing access to integrated databases. The process starts with configuring the Kerberos server and establishing trust relationships within the network to ensure reliable

access. The Kerberos authentication protocol is then implemented to verify user identities before granting access to the system. PHP's Kerberos extension is used to handle authentication requests and validate user credentials against the Kerberos Key Distribution Center (KDC). Security is further enhanced by encrypting communication channels between the client, server, and KDC using encryption algorithms supported by Kerberos.

DISCUSSIONS AND RESULTS

The proposed work has been secured using the Kerberos authentication algorithm. The analysis involved evaluating query retrieval times across 10 cases. This testing was performed on a 1.19 GHz Intel® Core™ i5-1035G1 CPU with 8GB of RAM running Microsoft Windows 10 20H2. All functionalities of the web application were developed using Visual Studio Code 6.0.

Figure 2 illustrates the input datasets sourced from four distinct databases: Banking, Qualification, Asset, and Aadhaar enrolment.



Fig. 2 : Database inputs

The output image shown in Figure 3 represents the verification page, which displays consolidated data from all the databases.



Fig. 3: Verification page

This succeeding table Table1 compares query performance before and after implementing indexing strategies.

Table 1: Indexing Performance Analysis

Query Type	Average Response Time Before Indexing (s)	Average Response Time After Indexing (s)	Improvement (%)
Aadhaar Lookup	1.2	0.3	75
Banking Details Retrieval	0.9	0.4	55
Asset Details Retrieval	1.0	0.5	50
Qualification Details View	0.8	0.3	62

This following table 2 compares the normalization performance before and after implementation.

Table 2: Normalization Efficiency

Normalization Type	Attributes before normalization (s)	Attributes before normalization (s)
4NF	Aadhar number in bank Aadhar number in aadhar enrollment Aadhar number in asset enrollment Aadhar number in qualification	Single attribute called aadhar number is used. (multi valued dependency)
1NF	Login and username can be duplication.	Duplication is removed.

The succeeding table 3 and table4 shows the efficiency of data retrieval by using the query retrieval time, considering all indexing and normalization measures.

Table3: Data Retrieval Efficiency

Data Type	Retrieval Method	Normalized	Indexed	Average Retrieval Time (s)
Aadhaar Details	Direct	Yes	Yes	0.3
Banking Details	Direct	Yes	Yes	0.4
Asset Details	Join	Yes	Yes	0.5
Qualification Details	Join	Yes	Yes	0.3

Table 4: Query Request Time

CASE NUMBER	QUERY RETRIEVAL TIME(milliseconds)
Case 1	0.019
Case 2	0.031
Case 3	0.022
Case 4	1.599
Case 5	0.571
Case 6	0.018
Case 7	0.007
Case 8	0.015
Case 9	0.013
Case 10	0.024
$\bar{T}$	0.2319

The resulting average for retrieving the data from the database($\bar{T}$) was calculated by the following formula:

$$\bar{T} = \frac{1}{n}\sum_{i=1}^{n} T_i$$

Where,

- $\bar{T}$ is the average retrieval time.
- Ti represents the retrieval time for each individual case, with i ranging from 1 to n.
- n is the total number of cases

This equation expresses the sum of all retrieval times (Ti) divided by the total number of cases (**n**), which yields the average retrieval time. By employing indexing techniques in the database, the retrieval time has been reduced to 0.23 milliseconds.

CONCLUSION

In conclusion, the "Aadhaar Integration Software" work represents a major step forward in data management and security. By consolidating information from Aadhaar enrollment, qualifications, banking, and asset databases, the system provides a unified platform for comprehensive data verification and processing. The implementation of Kerberos authentication ensures robust security measures, protecting sensitive information from unauthorized access. Through effective use of normalization and indexing techniques, the software enhances database performance, enabling rapid and accurate data retrieval.

This paper addresses the complex needs of various stakeholders, including government agencies and financial institutions, by streamlining data integration and ensuring compliance with regulatory standards. Ultimately, it improves operational efficiency and service delivery while maintaining the highest standards of privacy and data security.

REFERENCES

1. Abhishek Sarje, Shashikant Pattewar, Abhijit Pawar, Saurabh Nimse, " Document Storage and retrieval Using Adhaar Number"(2018) , Volume 3, Issue 1, pp. 4991-4993, 2018 ISSN 2395-1621.
2. Eman El-Emam, Magdy Koutb, HamdyKelash, and Osama Farag Allah, "An Optimized Kerberos authentication Protocol".

3. Pratik Joshi,Stevens, “Kerberos Security in distributed Systems”(2022),Volume 9, Issue 2, e-ISSN: 2395-0056 ,p-ISSN: 2395-0072.

4. Mahmoud T. El-Hadidi, Dr. Nadia H. Hegazi, Heba K. Aslan, “Performance Analysis of the Kerberos protocol in a Distributed Environment”.

5. Graham Bent, Dominic Harries, Andrew J. Toth, “Extending the Kerberos Protocol for Distributed Data as a Service ”(2012).

6. Justin Zobel, Hugh Williams, ”Trends in retrieval System performance”(2000), DOI:10.1109/ ACSC.2000.824410.

7. Fadi Al-Ayed, Chunqiang Hu and Hang Liu, “An Ecient Practice of Privacy Implementation: Kerberosand Markov Chain to Secure FileTransfer Sessions”(2017), DOI: 10.6633/IJNS.20180720(4).07.

8. B.Clifford Neuman, TheodoreTs’o, Kerberos: An Authentication Service for Computer Networks”(1994), Volume 32.

9. Robert M. Losee , Lewis Church Jr,” Information retrieval with Distributed Databases: Analytic models of Performance”(2003), Volume. 14, Issue 12.

10. Bryce Allen,” Perceptual Information Although Speed, Learning and Retrieval Performance”.

11. Nalin Subramanian, Chanjun Yang, and Wensheng zhang,” Securing Distributed Data Storage and retrieval in Sensor Networks”.

12. Juan A.Garayy, Rosario Gennaroz, charanjit Jutlaz, Tal Rabinz,” Secure Distributed storage and Retrieval”.

DOI - 10.68648/ijte-q-2026-49-03-014

ResNet50 Based Sign-to-Text-to-Speech Translation

Sheetal A. Takale
Information Technology Department
Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology
Baramati, Maharashtra
✉ sheetaltakale@gmail.com

Vaishnavi D. Jankar
Information Technology Department
Vidya Pratishthan's Kamalnayan Bajaj Institute of Engineering and Technology
Baramati, Maharashtra
✉ vaishnavijankar28@gmail.com

ABSTRACT

Today, effective communication is important in our connected world. However, for millions of deaf and mute people, daily communication is difficult. Sign language allows deaf and mute people to communicate with others. However, many people with normal hearing do not understand sign language. Existing automated sign language recognition systems often have low accuracy and perform poorly in different background conditions. This limits their use in real-world applications. This project aims to solve these problems. It develops a system that translates sign language into text and speech in real time. At its core, the system utilizes a custom CNN model and a pre-trained Convolutional Neural Network - ResNet50 to precisely detect and interpret manual gestures. Once the sign is recognized, it will be converted into text and then into speech. This research compares performance of custom CNN model and ResNet50 mode. The custom CNN model achieved accuracy of 88.2%. The ResNet50 model achieved a classification accuracy of 94%, demonstrating its effectiveness in sign language gesture recognition.

***KEYWORDS** : Sign Language Recognition, ResNet50, CNN, Gesture Recognition, Real-time Communication, Speech Conversion.*

INTRODUCTION

As per the World Health Organization (WHO) records, about 466 million people worldwide have hearing loss. This represents more than 5% of the global population. Hearing loss is the third most common physical disability in the world. Among those affected, about 34 million are children. Conventional communication between deaf and hearing individuals largely relies on human interpreters. However, availability of human interpreter is often limited. Also services of human interpreter can be costly and inconsistent in real-time environments. Around 72 million deaf individuals use sign language as their primary mode of communication, relying on more than 300 recognized sign languages worldwide. Sign languages used worldwide are listed in Table 1. However, less than 0.1% of people with normal hearing ability know sign language. This creates a major communication gap in the society. [1]. This results in an immediate need of automated sign language recognition and translation systems.

In the present research on sign language recognition and translation systems many of them use gloves with sensors. Issues to be addressed with such system are: cost effectiveness, discomfort resulting due to need of wearing it for long periods, and correctness of results [2]. This restricts their usability in everyday communication scenarios. A major This work is open access and licensed under the Creative Commons CC BY 4.0 License. Volume 12 (2022), Issue 6 limitation of many glove-based systems is need of static hand gestures. Their performance on dynamic gestures is limited. [3]. Existing vision-based glove-free approaches require controlled environments, including plain backgrounds, stable illumination, and fixed camera positioning, to achieve acceptable performance and improved accuracy [4]. Furthermore, several traditional and lightweight deep learning models exhibit high inference latency and limited classification accuracy, which constrains their applicability in real-time assistive communication systems [5].

The proposed system addresses these challenges using

a standard camera and the ResNet50 deep learning model unlike previous systems. It does not need special hardware, gloves, or a controlled environment [6]. The proposed solution is designed with simplicity in mind. User just has to perform natural sign language gestures. The system automatically recognizes the gestures. Deep learning model converts the recognized signs into text and speech. ResNet50 is a deep CNN model with residual learning. It is pretrained on ImageNet weights for transfer learning. It assures high accuracy and is well suited for sign language recognition [7]. It suitable for real-world applications becuase it is capable of handling variations in backgrounds and lighting conditions [8]. It assures accuracy for both static and dynamic hand signs in real-time. This ensures faster and more accurate results [9]. This project aims to improve the performance and practicality of sign language recognition in order the communication between deaf and hearing people. [10].

Table 1. Sign Languages by Region Summary

Region	Number of sign languages	Notable Examples
North America	10+	ASL, LSM, Maritime Sign Language
South America	15+	Libras, Argentinian SL, Colombian SL
Europe	30+	BSL, DGS, LSF, SSL
Asia	50+	JSL, CSL, ISL, KSL
Middle East	20+	Arabic SL, Israeli SL, Iranian SL
Africa	40+	SASL, Nigerian SL, Kenyan SL
Oceania	5+	Auslan, NZSL
Indigenous / Isolated	130+	Various village sign languages

LITERATURE REVIEW

Computer vision and deep learning has a major application for the problem of Sign language recognition (SLR) to develop a good communication between deaf and hearing people. Researchers have applied various technologies and methods for sign language recognition and translation including sensor-based, vision-based, and deep learning methods.

Preliminary approaches to Sign language recognition (SLR) use sensor-based systems to capture hand movements. Wearable devices like smart gloves equipped with flex sensors and accelerometers were used to detect finger positions and orientations. These approaches achieved an accuracy up to 96% for 20 letters of the alphabet.

Some systems use Arduino boards, flex sensors, and GSM modules to convert hand gestures into text and speech. Some use cloud-based learning to predict gestures. However, these systems require special hardware. This increases cost and reduces portability. They are also difficult to scale. As a result, vision-based methods have become more popular because they do not require specialized equipment.

Thakar et al. [10] proposed a model utilizing transfer learning with VGG16, achieving 98.7% accuracy. Their approach demonstrated high accuracy. However, it relied heavily on pre-trained models. This resulted in limitation on flexibility for custom datasets and new sign variations.

Duraisamy et al. [11] proposed a system where gesture recognition is done using convolution neural network, textual translation is done through NLP, and for speech synthesis neural TTS is used. Proposed approach achieved an accuracy of 97.6%. However, the use of multiple sequential deep learning components has resulted in high computation, slow model execution, and challenges in implementation on low resource devices.

For ASL finger spelling translation Munde et al. [12] developed a Deep Learning CNN model. Work carried out in this research particularly focuses on hand orientation and posture. Results obtained in this research show strong prediction performance. However, limitation of the system was limited performance to isolated static alphabet gestures. Proposed approach was unsuitable for continuous or dynamic sign sequences.

YOLOv5 for real-time sign recognition was proposed by Fathima et al. [13]. Proposed approach achieved Mean Average Precision (mAP) between 92% and 99% for 15 gesture classes.

Parashar et al. [4] compared performance of CNN model with ResNet50. This research reported that CNN models often miss minute gesture details. Where as Deep CNN ResNet50 can capture the minute details, enhancing recognition accuracy up to 88.7%. However, deeper networks increase computational load, and model struggled with complex hand gestures and background noise, making real-time deployment challenging on low-end devices.

Likewise, Barbhuiya et al. [7] used AlexNet and VGG16 architectures for Sign Language Recognition. This work succeeded with an accuracy of 99.82%. However, major concern was their experiments relied on relatively small datasets.

Stergiopoulou and Papamarkos [14] developed custon CNN model and achieved effective performance with 98% recognition accuracy. However, the approach proposed suffered due to requirement of high-performance hardware for high computational complexity.

Approach proposed by Haidar and Reefat [2] used sensor based gloves. It enabled highly accurate fine-grained gesture capture. However, the proposed approach suffered due to requirement of wearable hardware. It significantly reduced user comfort, portability, and scalability.

Approach proposed by Yan et al. [3] implemented Kinect-based depth sensing for dynamic sign recognition. It improved motion tracking accuracy. However, it suffered due to increase in overall system cost and hardware dependency.

In contrast, Singh et al. [15] utilized depth cameras to enhance gesture segmentation and to significantly improve the accuracy. Their approach enhanced gesture tracking but required expensive depth cameras.

Sequence modelling techniques is also used for dynamic gesture and temporal sign recognition. Abraham et al. [16] explored LSTM-based temporal modelling with glove sensors. It achieved a reliable dynamic gesture recognition but it requires a large sequential dataset, long training time, and significant computational resources. Similarly, Murakami and Taguchi [17] found that RNN-based models are computationally expensive and often struggle with unseen user patterns and signer variability. Xu et al. [18] implemented the S2VT sequence-to-sequence framework. However, the model suffered from high memory consumption, limiting its real-time applicability.

Garg et al. [19] proposed a deep learning-based sign language translator using advanced neural network architectures. Their system effectively converted sign gestures into text. However, their model was primarily focused on isolated gestures. It did not address dynamic gesture sequences or context-based sentence formation. This limits its usability for real-world conversations.

Survey of hand gesture recognition approaches proposed by Nandwana et al. [6] reported approaches which have accuracies exceeding 95%, however have issues regarding real-time deployment constraints, hardware optimization, or practical usability challenges. Mohamed et al. [1] identified that many existing SLR systems perform well on trained users howevre fail to generalize to new signers, dialectal variations, and uncontrolled environments. Table 2 presents a comparative analysis of recent deep learning-based sign language recognition studies published during 2025–2026. The comparison includes the deep learning model employed, dataset used, performance metrics such as accuracy, precision, recall, and F1-score (where reported), major findings, and identified limitations.

METHODOLOGY

Dataset

A dataset of American Sign Language (ASL) downloaded from Kaggle has image samples of of hand gestures with wide variety which are commonly used in ASL communication. The dataset has hand gestures for English alphabets, numbers, and additional commonly used symbolic gestures. It is organized into 43 distinct class folders, with each folder containing a unique gesture category.

The training dataset has images with 224 × 224 pixels resolution with a total of 105,600. These images have compatibility with deep learning architectures such as ResNet50. As listed in table 3, the gesture classes include 26 classes for alphabet signs (A–Z), 10 classes for numbers or numerical signs (0–9), and several additional gesture categories which are used in practical sign language communication. Each class contains

approximately 2,400 images, providing a balanced distribution across categories and reducing class imbalance during training.

However, some of the images in the dataset which are at a lower resolution of 50 × 50 pixels, are resized during pre-processing to preserve uniformity in feature extraction.

The diversity and balanced class representation of this dataset make it important benchmark for training robust ASL gesture classification models, accurate sign interpretation. The dataset is available at Kaggle - American Sign Language .

Table 2. Overview of the ASL Dataset

Category	Classes	Images per Class
Alphabets (A-Z)	26	2,400
Numbers (0-9)	10	2,400
English Words	7	2,400
Total	43	105,600

Data Pre-processing and Image Processing

Data pre-processing is a key step for improving the overall performance and accurate prediction on unseen data for deep learning models. In the proposed system, the first pre-processing step is all input images are first resized to a fixed resolution of 224 × 224 pixels. Figure 8 shows the images before and after resizing. It ensures dimensional consistency and compatibility with the ResNet50 architecture. Using fixed input size enables efficient batch processing and uniform feature extraction across the entire dataset.

Next step in pre-processing is normalization, the pixel values are divided by 255. This normalizes the values from 0–255 to 0–1. This normalization step help stabilize the gradient updates. It also accelerates convergence during training, and improves the network's ability to learn discriminative visual features.

Third step of pre-processing is data augmentation. This technique is used to improve model robustness and reduce the risk of overfitting. The data augmentation techniques used in this research include rotation, random horizontal flip. These augmentation techniques increased the size of dataset and diversity of the training data. It also improves model performance even under different hand positions, lighting conditions, backgrounds, and poses.

After pre-processing, for training and evaluation of the model, the dataset is divided into 80% training and 20% validation sets.

The proposed system architecture, shown in Figure 2, is designed as a complete real-time sign language recognition (SLR) and translation to text and speech framework. The process begins with the camera input module, where real-time video frames containing hand gestures are captured using a standard camera.

The captured frames are further pre-processed. Pre-processing involves operations such as resizing of images, color space conversion, noise filtering using Gaussian blur and thresholding, contour extraction, and normalization. These steps first isolate the hand region from the background, then reduce noise and effects of lighting variations.



Fig 1. Comparison of images before and after resizing (224x224)

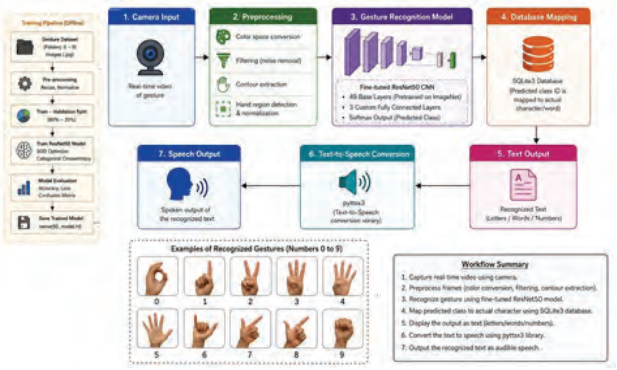


Fig 2. System Architecture

The pre-processed gesture images are given as input to the fine-tuned ResNet50 model with residual connections. It uses 49 pre-trained layers and 3 custom

classification layers to extract features and classify hand gestures. The residual connections help the model train faster and improve recognition accuracy.

The predicted gesture is matched with its corresponding letter, word, or number using an SQLite3 database. The database acts as a lookup table and converts the prediction into corresponding text. This design makes it easy to add new gestures in the future.

The mapped result is the text of the recognized letter, word, or number. The text is then converted into speech using the pyttsx3 text-to-speech (TTS) library. This allows users to hear the recognized sign and improves accessibility.

Hyper parameter such as the learning rate, batch size, number of epochs, and optimizer are adjusted to improve model performance. Dropout layers are added to reduce overfitting and improve the model's performance on new gesture images.

Proposed approach combines the steps of video acquisition, image pre-processing, ResNet50-based deep feature extraction and SLR classification, database-assisted semantic mapping, and speech synthesis. This proposed architecture provides an efficient and practical assistive communication system capable of functioning in diverse real-world environments.

Algorithm 1 Real-Time Sign Language Recognition and Speech Conversion

```
Initialize:
    Initialize webcam: cap = cv2.VideoCapture(0)
    Load model: model = load_model('resnet50.h5')
Define Preprocessing Function:
    def preprocess(frame):
        frame = cv2.resize(frame, (224, 224))
        frame = np.expand_dims(frame, axis=0)
        frame = frame / 255.0
        return frame
Start Video Capture Loop:
    while True:
        ret, frame = cap.read()
        if not ret: break
        roi = preprocess(frame)
        pred = model.predict(roi)
Convert Prediction to Text:
    label = decode_prediction(pred)
    print("Detected Sign:", label)
Convert Text to Speech:
    if voice_enabled:
        text_to_speech(label)
Display Output:
    cv2.imshow("Frame", frame)
    if cv2.waitKey(1) & 0xFF == ord('q'): break
Release Resources:
    cap.release()
    cv2.destroyAllWindows()
```

Model Architecture and Training

The proposed sign language recognition system is built upon the ResNet50 architecture as shown in table 4, a deep residual network pre-trained on the ImageNet dataset. To adapt the model for sign language classification, the top layers of the base model are replaced with a custom classification head. Specifically, a Global Average Pooling layer is applied to the output of the base model to reduce spatial dimensions, followed by a fully connected Dense layer consisting of 128 neurons with ReLU activation to learn high-level feature representations. A Dropout layer with a rate of 0.5 is incorporated to mitigate overfitting during training. Finally, a Softmax output layer is added, with the number of neurons corresponding to the total number of sign language classes, enabling multi-class probability estimation.

Table 3. ResNet50 Architecture Configuration

#	Layer	Configuration	Output Shape
1	Input Layer	224 × 224 × 3	(224, 224, 3)
2	ResNet50 Base	Pre-trained, Frozen, ImageNet weights	(7, 7, 2048)
3	Global Average Pooling 2D	Spatial averaging across feature maps	(2048,)
4	Dense Layer	128 units, ReLU activation	(128,)
5	Dropout Layer	Rate = 0.5	(128,)
6	Output Dense Layer	N units, Softmax activation	(N,)

Table 5. Hyperparameters Used for ResNet50 Model Training

#	Hyper-parameter	Value	#	Hyper-parameter	Value
1	Input Image Size	224 × 224 × 3	2	Batch Size	32
3	Number of Epochs	20	4	Validation Split	20% (0.2)
5	Optimizer	SGD	6	Learning Rate	0.001 (1e-3)
7	Momentum	0.9	8	Loss Function	Categorical Cross-Entropy
9	Evaluation Metric	Accuracy	10	Dropout Rate	0.5
11	Dense Layer Units	128	12	Activation (Hidden)	ReLU
13	Activation (Output)	Softmax	14	Pre-trained Weights	ImageNet
15	Base Model Layers	Frozen	16	Pooling Layer	Global Avg Pooling 2D

To leverage the rich feature representations learned from ImageNet, all layers of the ResNet50 base model are frozen during the initial training phase, ensuring that only the newly added classification layers are updated. Hyper parameters are listed in table5. Upon completion of training, the model's learning behavior is visualized through accuracy and loss curves plotted across epochs for both the training and validation sets. Upon successful completion of training, the optimized model is saved for subsequent deployment in the real-time sign language recognition pipeline.

Algorithm 2 Sign Language Gesture Recognition Training Using ResNet50

Input: Gesture image dataset
Output: Trained ResNet50 model
Data Preparation:
  Load gesture images and corresponding labels
  Resize images to (25×25) and normalize pixel values
  Encode labels into numerical format
  Split dataset into training and validation sets $(80\% : 20\%)$
  Store preprocessed dataset
Model Construction:
  Load pretrained ResNet50 without top layers
  Add Global Average Pooling layer
  Add Fully Connected layer with ReLU activation
  Add Dropout layer with $p = 0.5$
  Add Softmax output layer for classification
  Compile model using SGD optimizer and categorical cross-entropy loss
Training:
  Train model for $EPOCHS = 2$ with batch size $= 32$
  Save best model based on validation accuracy
Evaluation:
  Plot training and validation accuracy/loss curves
  Compute confusion matrix for performance analysis
Return: Trained model and evaluation metrics

Text-to-Speech Conversion

The system includes a Text-to-Speech (TTS) module. After a gesture is recognized, the predicted letter, or word, is displayed as text. The text is then sent to the TTS engine. The pyttsx3 speech synthesis library converts the text into speech. This allows hearing people to understand the recognized sign. By providing both text and speech output, the system improves communication between deaf and hearing individuals.

Convolution Neural Network Model Architecture and Training

The proposed sign language recognition system uses a custom designed Convolutional Neural Network (CNN) to classify hand gestures. The CNN architecture, shown in Figure 8, performs feature extraction and classification of gestures. The input to the CNN model is a pre-processed 256 × 256 × 1 grayscale gesture image. The first three convolution layers are used to extract features from gesture image. The first convolution layer applies 16 filters of size 2 × 2 with ReLU activation, followed by 2 × 2 max pooling. The second convolution layer uses 32 filters of size 5 × 5, followed by 5 × 5 max pooling. A third convolution layer with 64 filters of size 5 × 5 are further used. The resulting feature maps are then flattened into a one-dimensional feature vector and passed to a fully connected (dense) layer with 128 neurons using the ReLU activation function. A dropout layer with dropout rate of 0.2 is employed to reduce overfitting and improve the model's generalization capability. Finally, the output layer consists of N neurons, where N represents the 43 different gesture classes. Output layer uses the Softmax activation function to predict the probability of each of the 43 classes. The network is trained using the Cross-Entropy loss function with the Gradient Descent optimizer with the learning rate of 0.01. This architecture effectively extracts spatial features from hand gesture images and performs accurate multi-class classification for sign language recognition.

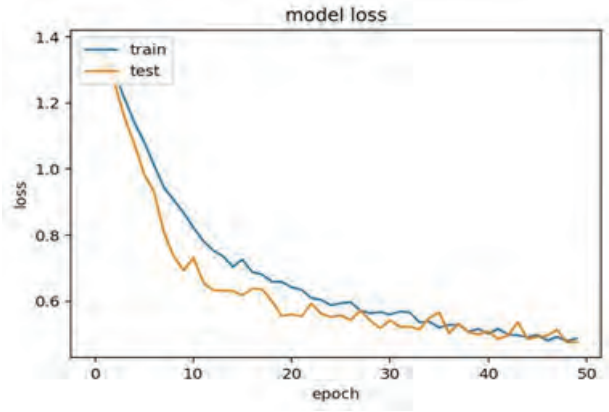


Fig 3. Training Results

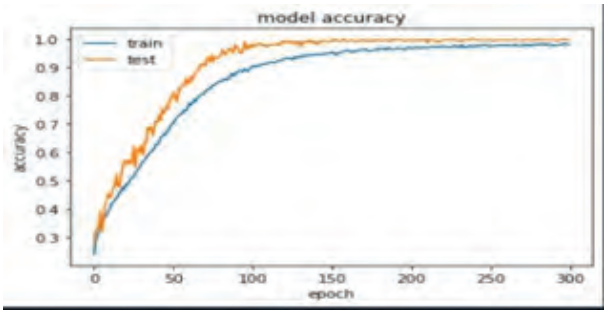


Fig 4. Training Results

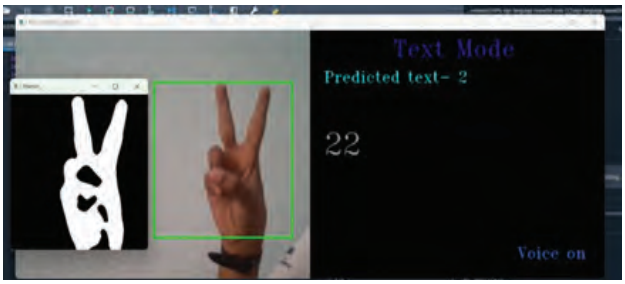


Fig 5. Output

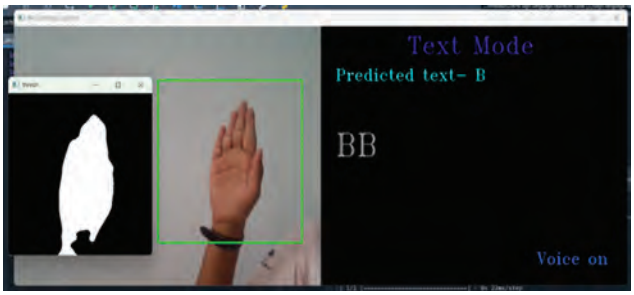


Fig 6. Output

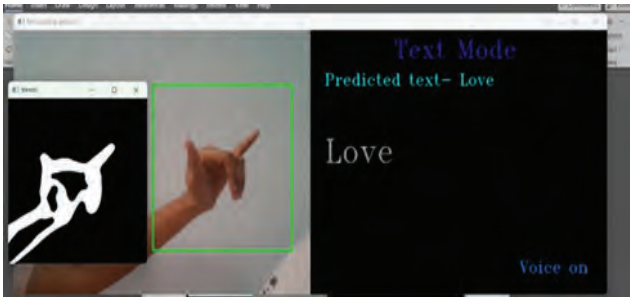


Fig 7. Output

An 80:20 train-validation split is used in this implementation to ensure balanced distribution. Data augmentation techniques such as rotation, flipping, contrast adjustment improve performance and generalization ability of the model for the real-world variations. The network is trained using the categorical cross-entropy loss function and optimized using the Adam optimizer with a learning rate of 0.001, ensuring stable and fast convergence. Mini-batch gradient descent is employed with a batch size of 32, and the model is trained for 10 epochs. To further improve robustness, early stopping monitors validation accuracy and halts training when no significant performance gain is observed. The model weights are initialized using He initialization, which facilitates stable gradient propagation across deeper layers.

RESULTS

Experimental Setup

The two deep learning models : ResNet50 and custom CNN model are implemented and trained on a high-performance AI workstation featuring a 14th Gen Intel® Core™ i9-14900K processor with 64 GB DDR5 RAM, and an NVIDIA® GeForce RTX™ 4090 GPU with 24 GB GDDR6X memory. The powerful hardware configuration enabled reduced the training time.



Fig 8. Convolutional Neural Network Model Architecture

Algorithm 3 Real-Time Gesture Recognition Using Convolutional Neural Network

```
Load the Model:
    Load pretrained CNN model for gesture classification
Initialize Video Capture:
    Start device camera for real-time video input
Define Detection Box:
    Draw rectangular region for hand placement
Extract Region of Interest (ROI):
    Capture image inside detection box
Preprocess ROI:
  ROI = resize(ROI, (64, 64))
  ROI = normalize(ROI)
  ROI = reshape(ROI)
Make Predictions:
  prediction, confidence = model.predict(ROI)
Confidence Filtering:
  if confidence > 0.9:
    accept_prediction()
  else:
    discard_prediction()
Display and Convert Gesture:
    Display recognized letter on screen
    text_to_speech(prediction)
Exit and Release Resources:
    Stop video capture
    Release system resources
```

Output

Figure 3 and Figure 4 illustrates the ResNet50 model's loss and accuracy over 50 epochs for both training and testing datasets. Figure 5, 6 and 7 show the system output for ResNet50 model, where the hand gesture is detected and classified in real time. The detected hand gesture is

shown in closed green bounding box. The recognized letter is displayed on the screen. GUI guides the user to position their hand correctly for accurate detection. The predicted text is displayed on the screen, and the voice output is also generated, making the system interactive and accessible.

Figure 12 demonstrates model training log of the proposed CNN model. Figure 9 and Figure 10 illustrates the custom CNN model's loss and accuracy results. Figure 11 show the system output for CNN model, where the hand gesture is detected and classified in real time.

Comparison of Model Performance

Table 6. Performance Results

Model	Accuracy (%)
ResNet50	94.5
CNN	88.2

Table 6 shows a comparative study of different algorithms for sign language recognition. Among the models, ResNet50 gives a high accuracy (94.5%) with lower complexity compared to the deep learning model with custom CNN. ResNet50 is preferred for ef4.4. Evaluation Metrics

In sign language recognition, evaluating model performance is essential to ensure accurate translation of gestures into text. The most commonly used metrics are Precision, Recall, Accuracy, and F1-Score.

Figure 13 illustrates the confusion matrix of the proposed CNN model for recognizing 43 sign language classes. The matrix exhibits a strong diagonal pattern, indicating that the majority of test samples are correctly classified into their respective classes. The dark blue diagonal cells represent high true positive predictions.

CONCLUSION

This paper presented a sign language recognition system that converts static hand gestures for 43 different classes including alphabets, numbers and few commonly used symbolic gestures into text and speech using deep learning. Two models, a custom CNN and the pre-trained ResNet50, were implemented and evaluated. Experimental results showed that the custom CNN achieved an accuracy of 88.2%, while ResNet50 achieved 94% accuracy. The higher performance of ResNet50 demonstrates the advantage of transfer learning for extracting meaningful features from hand gesture images. By providing real-time text and speech output, the system has the potential to improve communication between deaf or mute individuals and people who are not familiar with sign language. Future improvements include integrating deep learning for dynamic recognition, optimizing real-time processing, and incorporating facial expressions and body posture for enhanced accuracy and usability, efficient and accurate performance.

Table 7. Performance Evaluation Metrics

Metric	Definition	Formula
Accuracy	Percentage of correctly classified samples among all samples.	$\frac{TP+TN}{TP+TN+FP+FN}$
Precision	Percentage of predicted positive samples that are actually positive.	$\frac{TP}{TP+FP}$
Recall	Percentage of actual positive samples correctly identified by the model.	$\frac{TP}{TP+FN}$
Confusion Matrix	A table that summarizes the classification results using true and predicted classes.	TP - True Positive , TN- True Negative, FP- False Positive, FN- False Negative.

Table 8. Confusion Matrix

Actual	Predicted	
	Positive	Negative
Positive	TP	FN
Negative	FP	TN

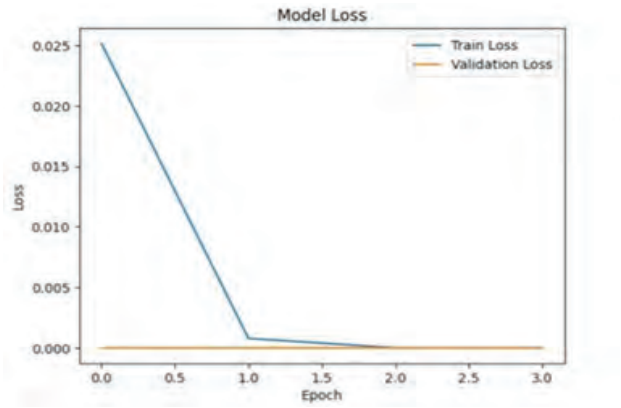


Fig 10. Training Results

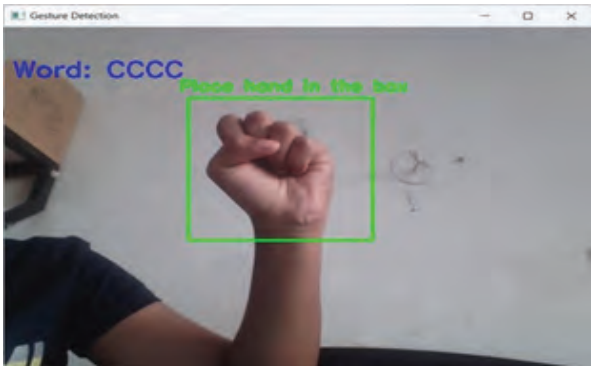


Fig 11. Experimental Result

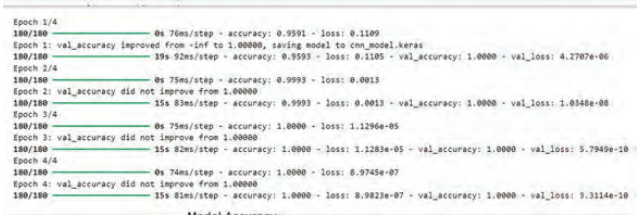


Fig 12. Training Epoch

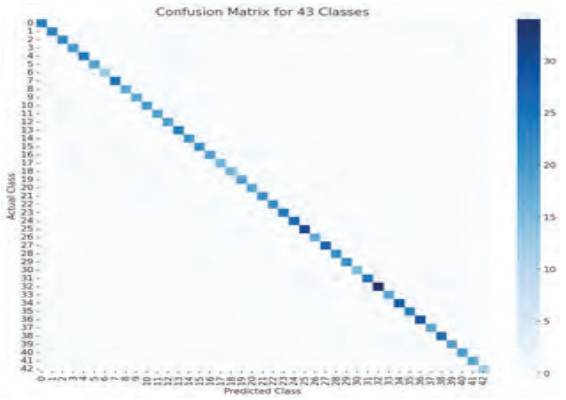


Fig 13. Confusion Matrix

REFERENCES

1. N. M. et al., "A review of the hand gesture recognition system: Current progress and future directions," IEEE Access, vol. 9, pp. 157422–157433, 2021.
2. G. I. Haidar and H. I. Reefat, "Glove-based american sign language interpretation using cnn and data glass," in Proc. IEEE Int. Conf. on Artificial Intelligence, 2020.
3. Y. Y. et al., "Research on dynamic sign language algorithm based on sign language trajectory and key frame extraction," IEEE Trans. Image Process., vol. 28, no. 3, pp. 1–10, 2019.
4. D. P. et al., "A deep learning-based approach for hand sign recognition using cnn architecture," IEEE Access, vol. 9, pp. 115322–115335, 2021.
5. H. S. L. Vaishnavi D. A., A. Kumar C. and D. M. L., "Mediapipe to recognise the hand gestures," WSEAS Trans. Signal Process., vol. 18, pp. 134–140, 2022.
6. B. N. et al., "A survey paper on hand gesture recognition," Int. J. Comput. Vision Signal Process., vol. 8, no. 2, pp. 45–55, 2020.
7. A. A. B. et al., "Cnn-based feature extraction and classification for sign language," in Proc. IEEE Int. Conf. on Computer Vision, 2021.
8. S. T. et al., "Sign language recognition and converting into text," J. Comput. Sci. Appl., vol. 12, no. 4, pp. 67–75, 2019.
9. Y. Y. et al., "Video hand gesture recognition using depth camera and lightweight cnn," in Proc. IEEE Int. Conf. on Artificial Intelligence and Machine Learning, 2022.
10. S. T. et al., "Sign language to text conversion in real-time using transfer learning," in Proc. IEEE Int. Conf. on Intelligent Computing and Control Systems (ICICCS), 2021, pp. 345–350.
11. P. Duraisamy et al., "Transforming sign language into text and speech through deep learning technologies," SRS Journal, December 2023, accessed 2024.
12. M. Munde et al., "A real-time sign language to text conversion system for enhanced communication accessibility," Journals.tultech.eu, August 2024, accessed 2024
13. T. Fathima et al., "Real-time sign language recognition and translation using deep learning techniques," Irjaeh, 2024, accessed 2024.
14. E. Stergiopoulou and N. Papamarkos, "A new technique

for hand gesture recognition (sgong method)," Pattern Recognition Letters, vol. 42, pp. 1–9, 2019

15. M. K. S. et al., "Hand gesture recognition with depth images: A review," in Proc. IEEE Int. Conf. on Image Processing and Machine Vision, 2020
16. E. A. et al., "Real-time translation of indian sign language using lstm," in Proc. Int. Conf. on Neural Networks and Signal Processing, 2020
17. K. Murakami and H. Taguchi, "Gesture recognition using recurrent neural networks," Fujitsu Laboratories Ltd., Kawasaki, Japan, Tech. Rep., 2020.
18. B. X. et al., "Application of tensor train decomposition in s2vt model for sign language recognition," IEEE Trans. Signal Process., vol. 68, pp. 1124–1135, 2021.
19. H. Garg et al., "Deep learning-based sign language translator," in Springer Lecture Notes, July 2024, accessed 2024.
20. A. Almjally and W. S. Almukadi, "Deep computer vision with artificial intelligence based sign language recognition to assist hearing and speech-impaired individuals," Scientific Reports, vol. 15, no. 1, pp. 1–15, 2025.
21. M. Balat, R. Awaad, A. B. Zaky, and S. A. Aly, "Revolutionizing communication with deep learning and xai for enhanced arabic sign language recognition," arXiv preprint arXiv:2501.08169, 2025.
22. B. Alabduallah and A. A. Alneil, "Revolutionizing sign language recognition for hearing-impaired persons using ensemble of deep learning techniques with fine tuning model," Journal of King Saud University - Computer and Information Sciences, vol. 38, no. 2, 2026. [Online]. Available: https://link.springer.com/article/10.1007/ s44443-025-00395-7
23. A. I. Rastogi, A. I. Mahapatra et al., "Advanced gesture recognition in Indian Sign Language using a synergistic combination of YOLOv10 with Swin Transformer model," Scientific Reports, vol. 15, p. 32894, 2025. [Online]. Available: https://www.nature.com/articles/s41598-025-18496-8
24. J. Hao and H. Pan, "A novel deep transformer based cvt model for sign language recognition in visual communication," Scientific Reports, vol. 16, pp. 1–15, 2025.
25. H. Harrouch, L. Trabelsi, M. Jebali, and O. Gammoudi, "A deep learning-based method combines manual and non-manual features for sign language recognition," Scientific Reports, vol. 16, no. 1, p. 2981, 2025.
26. V. K. Gajjala, N. T. J., L. R. Battarusetty, and V. Avula, "A vision-based sign language recognition & voice conversion system for inclusive communication," Scientific Reports, vol. 16, no. 1, p. 35541, 2026. [Online]. Available: https://www.nature.com/articles/s41598-026-35541-2
27. J. K. Nath, S. A. M. Alam, Y. Ullah, and Z. H. Abdullah, "Visual deep learning approaches for alphabetic sign language interpretation," Scientific Reports, vol. 16, no. 1, p. 57945, June 2026.
28. R. Gupta and S. R. Nalla, "Efficientsign: An attention-enhanced lightweight architecture for indian sign language recognition," 2026. [Online]. Available: https://arxiv.org/abs/2604.08694

Table 2. Comparison of Recent Deep Learning-Based Sign Language Recognition Methods (2025–2026)

Research Paper	Model	Dataset	Acc.	Prec.	Rec.	F1	Limitations
Almjally et al. (2025) [20]	ResNet-152 + BiLSTM + Harris Hawk Optimization	Kaggle Sign Language for Numbers (15,000 images, 10 classes)	98.95	94.72	94.74	94.72	Evaluated only on static number gestures; continuous recognition and text-to-speech translation were not considered.
Mazen Balat et al. (2025) [21]	ResNet50, MobileNetV3, EfficientNet-B2 + XAI	ArSL2018, AASL	99.48	NR	NR	NR	Limited to Arabic Sign Language datasets.

Alabduallah et al.(2026) [22]	InceptionV3 + BiLSTM + GRU + VAE + Tasmanian Devil Optimization	American Sign Language Dataset	98.18	76.77	76.52	76.45	High computational complexity; evaluated only for isolated ASL gestures.
Rastogi et al. (2025) [23]	YOLOv10 + Swin Transformer + Mish Activation	Custom Indian Sign Language Dataset (15,000 images + 35 videos)	97.62				
(mAP)	97.50	98.10	96.58	Small dynamic video dataset; focuses on gesture recognition rather than complete translation.			
Hao and Pan (2025) [24]	Convolutional Vision Transformer (CvT)	Sign Language Digits Dataset; ASL Alphabet Dataset	99.00	NR	NR	NR	Evaluated only on isolated digits and alphabet datasets.
Haifa Harrouch et al. (2026) [25]	Hybrid Deep Learning Model	Public Sign Language Dataset	NR	NR	NR	NR	Higher computational complexity due to multimodal feature extraction.
Gajjala et al. (2026) [26]	CNN + Transformer	Custom Gesture Dataset	99.00	NR	97.00	97.00	Limited vocabulary and language-specific evaluation.
Joyonta Kumar Nath et al. (2026) [27]	ConvNeXtXLarge, EfficientNet, VGG19, ResNet50	ASL Alphabet Dataset	99.81	NR	NR	NR	Limited to alphabet-level recognition.
Gupta et al. (2026) [28]	EfficientNet-B0 + Attention	Indian Sign Language Alphabet Dataset (12,637 images)	99.94	NR	NR	NR	Evaluated only on alphabet-level Indian Sign Language.

DOI - 10.68648/ijte-q-2026-49-03-015

Dynamic Wireless Electric Vehicle Charging with Seamless LCC-S Pad Handover Control

Suganthi. J.
Professor
Department of Electrical and Electronics Engineering
Government College of Engineering,
Tirunelveli, Tamil Nadu
✉ suganthi@gcetly.ac.in

ABSTRACT

The demand for efficient and user-friendly charging infrastructure has led to the growth of electric vehicles. Wireless power transfer technology offers a compelling alternative; however, existing dynamic wireless power transfer implementations often rely on complex vehicle positioning systems, inter-pad communication protocols, and real-time adaptive control algorithms, all of which contribute to system complexity and reduced reliability. This paper presents a simplified dynamic wireless charging architecture design for vehicles in motion, employing LCC-S compensation topology. In the proposed system, the transmitter pads are sequentially activated in a defined time domain frame and they are characterized with distinct mutual inductance values. A front-end AC-DC converter rectifies a 230V RMS supply to a 325V DC, which is subsequently elevated to 600V DC through a boost converter incorporating power factor correction. The LCC-S network with high-frequency inverter operating at 85-92 kHz is tuned to resonance so as to suppress reactive power. The simulation result confirms stable output characteristics, with the battery terminal voltage of 60V and an average current of 10A sustained throughout the pad transition intervals. The proposed system maintains an uninterrupted power supply without detectable discontinuity during pad handover events.

KEYWORDS *: Dynamic Wireless Power Transfer, LCC-S Compensation, CC-CV Charging, Sequential Pad Activation, Pad Handover Control.*

INTRODUCTION

Worldwide, the development of transportation systems is undergoing a fundamental shift. There is an urgency to reduce the carbon emissions, which, combined with falling battery costs and improving vehicle range, has pushed electric vehicles from an experimental category onto everyday roads. In many continents the governments have announced phased bans on internal combustion engine vehicles, and major automotive manufacturers have committed to all-electric lineups within the next decade. In practice, the main obstacle keeping it slow is charging. The way most electric vehicles are charged today is not fundamentally different from how a mobile phone is charged. We have to stop the vehicle and then plug in the charger, and then we have to wait until it is getting charged. For short urban commutes, this is manageable. But for highway travel, logistics fleets, or emergency vehicles, it becomes a real operational problem. A long truck cannot afford to sit idle for forty minutes every few hours. A city bus running a tight schedule cannot pull over for a charging stop. The top reason people hesitate before buying an EV is that for private car owners, the fear of running out of charge before reaching a station. Even if we increase the number of fixed charging stations, it helps, but it does not eliminate the fundamental limitation: the vehicle must stop [1].

Wireless power transfer technology offers a different starting point. Instead of a cable connecting the vehicle to a power source, energy is transferred across an air gap through magnetic coupling between a transmitter coil embedded in the ground and a receiver coil

mounted beneath the vehicle [2]. Stationary wireless charging, where the vehicle parks over a charging pad, has already reached commercial deployment in some markets. But the more ambitious version of this idea is dynamic wireless charging, where power transfer happens while the vehicle is moving [3]. If dynamic wireless power transfer is implemented successfully, it could allow an electric vehicle to charge continuously along a dedicated road segment, reducing the battery size needed for long trips and potentially eliminating range anxiety altogether. Dynamic wireless power transfer is not a new technology. Since the 2010 many research groups have been exploring it, and several road trials have been conducted in South Korea, Sweden, Germany, and the United Kingdom. What those projects have consistently revealed, however, is that the control side of the system is surprisingly difficult to get right. If the vehicle moves over a series of transmitter pads then the coupling between the active pad and the vehicle receiver changes continuously. When the vehicle approaches the pad directly overhead, the mutual inductance rises and falls as it moves away. To manage this charging coupling without disrupting the power delivery to the battery requires either very fast real-time control or some form of coordination between pads to ensure a smooth handover.

The major issue faced in this dynamic WPT, however, is more nuanced than simply embedding coils in a road [4]. The magnetic coupling between the active transmitter and the vehicle receiver charges continuously, when a vehicle moves over a series of transmitter pads. The power delivered to the battery varies when there is a change in coupling variation, and it is a major issue. If the charging system fails to respond properly, variations in output voltage and current can occur, which may disturb the charging process and prevent the battery from receiving a stable and controlled charge. Modern EVs require a specific two-stage profile used for battery charging, particularly for lithium-based cells. In the first stage, the vehicle charges in Constant Current (CC) charging mode regardless of the battery's terminal voltage, allowing the cell to absorb energy rapidly. In the second stage, when the voltage reaches a defined threshold, the system shifts to Constant Voltage (CV) charging, where the voltage is held constant while the current gradually tapers. This CC-CV profile is not simply a performance preference – it is a fundamental requirement for battery safety and longevity [5]. Deviating from it, even temporarily, can accelerate cell degradation or trigger protective shutdowns. It is harder in a dynamic wireless system to maintain a CC-CV charging profile than a stationary one. The power delivered through the coils varies continuously. When the vehicle moves across the pads and without active compensation, the output seen by the battery will not follow the required profile. Most existing dynamic WPT research has focused on achieving power transfer continuity during pad transitions, but far fewer studies have explicitly addressed the problem of maintaining a well-regulated CC-CV profile throughout the entire period of vehicle movement [6].

In existing approaches the output regulation typically relies on vehicle position feedback to anticipate coupling changes and adjust inverter parameters accordingly. Others use communication between pads to coordinate switching and prevent output drops during handover. These strategies can achieve reasonable regulation, but they introduce system complexity that works against practical deployment. Position sensing hardware must be installed and maintained. In addition to this, the communication between system components should be fast and dependable, and the control mechanism needs to be adjusted according to factors like vehicle speed and the spacing between transmitter pads. Each of these dependencies increases cost and reduces robustness [7].

The work presented in this paper takes a fundamentally different approach. The objective is to maintain a stable CC-CV charging output at the battery terminals throughout the vehicle's passage over multi-pad communication. To achieve this, an LCC-S compensation topology is employed on this transmitter side. The LCC-S network is known for its ability to present a near-constant current source characteristic to the receiver, making it inherently more tolerant of coupling variation than simpler series or parallel compensation schemes [8]. For simulation purposes, vehicle movement is represented through a time-domain model in which three transmitter pads with distinct mutual inductance values are activated sequentially across defined time intervals. This approach replicates the coupling variation of a moving vehicle. In hardware implementation the time-based switching would be

replaced by position-triggered activation, but the compensation network and output regulation principle remain the same [9].

SYSTEM OVERVIEW

The proposed dynamic wireless EV charging system is built with the layered architecture that separates the power conditioning functions from the wireless energy transfer functions as shown in Fig. 1. The overall system easier to analyze and validate through simulation, and hence, this separation is intentional, and it allows each stage to be designed and optimized independently. The signal flows from the AC mains supply on the ground side through to the battery terminals on the vehicle side as shown in the block diagram. The block diagram represents the complete architecture of dynamic wireless power transfer [10, 11].

The ground-level infrastructure can be understood in two parts. The first part is the shared power conditioning chain, which takes energy from the grid and prepares it for high-frequency transmission. The second part is the pad array, where three independent transmitter units are arranged along the charging lane, each capable of coupling energy into the vehicle receiver as the vehicle passes overhead. These two parts are connected in a common bus arrangement, meaning all three pads draw from the same rectified and boosted DC supply.

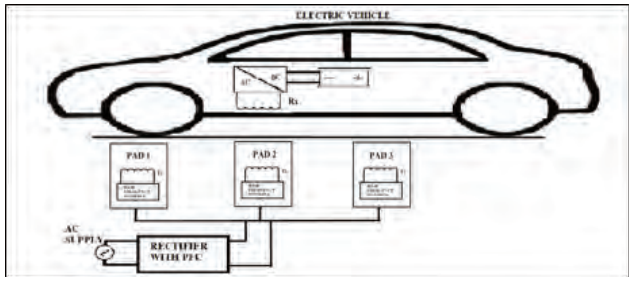


Fig. 1: Block Diagram of Proposed Dynamic Wireless Power Transfer

At any instant, only one transmitter pad is active, and its operation is governed by a simple sequential switching approach. This switching logic decides which pad should be energized based on the vehicle's movement along the charging path, allowing the system to transfer power in a controlled and continuous manner. The proposed system runs on the standard 230V AC power supply that is available for our home or office, which means the system, can be plugged in with the existing electrical setups, and there is no need for any extra wiring or high voltage requirement. Initially the AC power supply passes through a rectifier, which converts the alternating current into a more usable DC form with the peak values of 325V. However, this DC is not perfectly smooth. It fluctuates, and the path of power from the supply can cause electrical noise and inefficiencies that may affect other devices on the same network [12].

To solve both of these issues, a power factor correction stage is integrated in the system. The PFC circuit, implemented as a boost-type converter operating in continuous conduction mode, serves a dual purpose. First, it shapes the input current waveform to follow the sinusoidal voltage waveform, bringing the power factor close to unity and reducing harmonic injection into the grid. The rectified DC is boosted and maintained at a steady state at around 600V. . This stable voltage supplies the rest of the system, protecting it from any fluctuations or changes while the vehicle is drawing the power at any given moment. Each of the three charging pads has its own dedicated inverter, essentially its own power controller. This inverter takes that steady DC voltage and converts it into a rapidly alternating AC signal, cycling between 85 and 92 kHz. This high-frequency signal is then passed through a specially designed filtering network before reaching the transmitter coil, where it's ready to wirelessly transfer power to the vehicle. Operating at this frequency range is a deliberate choice; higher frequencies allow smaller coil geometries and compensation components for a given power level, and the 85 to 92 kHz band falls within the range recommended by international standards for inductive power transfer systems, specifically SAE J2954 and IEC 61980.

The three transmitter pads labeled Pad 1, Pad 2, and Pad 3 in the block diagram are physically spaced along the road surface beneath the vehicle travel path. Each pad consists of a transmitter coil wound to a specified inductance, paired with an LCC compensation network tuned to the operating frequency. The mutual inductance between each transmitter and the vehicle receiver varies depending on the vehicle's lateral and longitudinal position relative to that pad [13]. In this study, the

mutual inductance values are distinct for three pads. A higher mutual inductance value is assigned to Pad 2, as this represents the condition when the vehicle is directly aligned with the transmitter pad, resulting in stronger coupling. In contrast, Pads 1 and 3 are assigned lower values to represent the reduced coupling that occurs when the vehicle is entering or leaving the charging region. This asymmetric but realistic distribution of coupling values allows the simulation to test the system's ability to maintain CC-CV output across genuinely different operating conditions, not just a uniform worst-case scenario.

On the vehicle side, a single receiver coil mounted beneath the chassis picks up the magnetic flux generated by whichever transmitter pad is currently active. The receiver coil is part of the series compensation network on the secondary side of the LCC-S topology, which resonates at the same operating frequency as the transmitter. Maintaining resonance on the receiver side allows the induced voltage to be transferred more effectively to the output stage. This reduces unnecessary reactive power circulation and improves the overall efficiency of the power transfer process. Once the wireless power reaches the vehicle's receiver coil, it's still in AC form, so it gets converted back into DC on the spot before being sent to the battery.

In the proposed work it is designed to charge the battery at a nominal voltage of 60V, which keeps it practical for simulation and testing purposes. This charging is maintained constant by maintaining a steady current of around 10A, and the voltage is maintained at 60V to avoid overcharging. Upholding this profile continuously is the primary performance objective of the entire system even as the active pad changes and the mutual inductance shifts. The block diagram captures this end-to-end flow clearly, from the grid connection at the bottom left to the battery terminals at the top right, with the wireless air gap between the road surface and the vehicle underside forming the critical link in the middle [14].

SYSTEM DESIGN

The proposed dynamic wireless EV charging system consists of three interconnected design stages: a boost converter with power factor correction, an LCC-S compensation network, and a multi-pad transmitter array with defined coupling characteristics [15]. Each stage is designed independently and integrated into a unified power chain that delivers a regulated CC-CV output to the vehicle battery. The design parameters for each stage are presented in the following subsections.

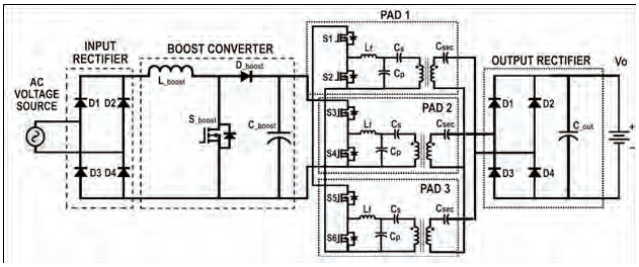


Fig. 2 Circuit Diagram

Boost Converter Design

The boost converter's job is to take the rectified voltage and push it up to a higher, stable DC level that the inverter can reliably work with. Duty cycle determines how long the switch stays in the on condition as indicated by equation (1). The voltage conversion ratio is governed by:

$$\frac{Vout}{Vin} = \frac{1}{(1-D)} \quad (1)$$

To determine the inductor value L, the switching frequency fsw and the allowable current ripple Δi_L are the two primary factors considered. A higher switching frequency naturally reduces the required inductance, while a tighter ripple limit demands a larger inductor. This relationship is captured as in equation(2):

$$L = \frac{(Vin \times D)}{(fsw \times \Delta iL)} \quad (2)$$

The selection of the capacitor is indicated in Equation (3) is done based on the output voltage ripple ΔVout to an acceptable level:

$$C = \frac{(Iout \times D)}{(fsw \times \Delta Vout)} \quad (3)$$

Where Iout is the average output current and ΔVout is the peak-to-peak voltage ripple.

A larger capacitor is selected beyond the minimum requirement to attenuate voltage disturbances caused by pad switching transients on the load side. The complete

set of boost converter design parameters is listed in Table 1.

Table. 1. Boost Converter Design Parameters

Parameter	Symbol	Value
Duty Cycle	D	0.45
Input Voltage	Vin	325 V
Output Voltage	Vout	600 V
Inductor Current Ripple	ΔiL	20%
Output Voltage Ripple	ΔVout	5%
Boost Inductor	L	0.8 mH
Output Capacitor	C	680 µF
Switching Frequency	fsw	65kHz

LCC-S Compensation Network Design

The LCC-S topology is employed on the transmitter side to achieve a near-constant current source characteristic at the output of the transmitter coil [16]. This behaviour arises from the resonance condition of the series branch formed by the filter inductor Lf and the series capacitor Cs, which forces the transmitter coil current to be determined primarily by the inverter voltage and Lf rather than by the reflected impedance from the receiver. As a result, moderate changes in mutual inductance as encountered when the vehicle moves between pad positions do not significantly disturb the transmitter coil current, reducing the burden placed on the output regulation stage [17].

The resonance condition for the transmitter series branch is indicated by the equation(4).

$$\omega_1 = \frac{1}{\sqrt{(Lf \cdot Cs)}} \quad (4)$$

where ω = 2πf is the angular operating frequency. To determine the value of parallel capacitor Cp, is indicated in equation(5), it is selected to resonate with the transmitter coil inductance L at the same frequency:

$$\omega_1 = \frac{1}{\sqrt{(L_1 \cdot Cp)}} \quad (5)$$

On the receiver side, the series compensation capacitor Csec is tuned as indicated in equation (6) to resonate with the receiver coil inductance L_2 at ω_1:

$$\omega_1 = \frac{1}{\sqrt{(L_2 \cdot Csec)}} \quad (6)$$

At resonance, the transmitter coil current, ITx, is given by:

$$ITx = \frac{Vinv}{(j\omega_1 \cdot Lf)} \quad (7)$$

Where, Vinv is the fundamental component of the inverter output voltage. Equation (7) confirms that ITx is independent of the load reflected from the receiver, which is the key property exploited in this design to maintain stable power delivery across varying coupling conditions [18]. The LCC-S network parameters are summarized in Table 2.

Table. 2. LCC-S Compensation Network Parameters

Parameter	Symbol	Value	Unit
Series Filter Inductor	Lf	60	µH
Series Capacitor (Tx side)	Cs	58	nF
Parallel Capacitor (Tx side)	Cp	150	nF
Series Capacitor			
(Rx side)	Csec	99	nF
Transmitter Coil Resistance	Rp	0.2	Ω
Receiver Coil Resistance	Rs	0.1	Ω
Operating Frequency	f_1	85	kHz

Mutual Inductance and Coupling Coefficient

The magnetic coupling between each transmitter pad and the vehicle-side receiver is characterized by the values of mutual inductance M, as indicated by the equation (8) which is related to the transmitter coil inductance L_1, the receiver coil inductance L_2, and the coupling coefficient k by:

$$M = k \cdot \sqrt{(L_1 \cdot L_2)} \quad (8)$$

Table. 3. Coupling Coefficient Ranges for Dynamic Operation

Alignment Condition	Coupling Coefficient (k)	Physical Scenario
Perfect alignment	0.25 – 0.30	Receiver centered above Tx pad
Partial alignment	0.15 – 0.20	Receiver partially offset from pad

Weak coupling	0.05 – 0.10	Receiver at pad entry or exit region

The coupling coefficient k is a dimensionless quantity that ranges between 0 and 1 and is determined by the geometry, separation, and lateral alignment between the transmitter and receiver coils. In dynamic WPT applications, k varies continuously as the vehicle moves relative to each pad [19]. Three alignment conditions are defined for this study, as listed in Table 3.

In this study, vehicle movement along the charging lane is represented through a time-domain model in which each pad is assigned a fixed mutual inductance value corresponding to its coupling condition during the simulation interval. The design of three pads in the proposed system is such that Pad 2 is placed close to the centre alignment which is assigned with the highest mutual inductance [20]. The positions of pads 1 and 3 are assigned with weak mutual inductance to represent during the vehicle entry and exit phases, respectively. These values represent the partial alignment condition and reflect the reduced coupling at the periphery of the charging lane. The pad activation sequence and corresponding mutual inductance values are given in Table 4.

TABLE 4: Pad Activation Sequence and Mutual Inductance Values

Pad	Coupling Coefficient (k)	Mutual Inductance (M)	Condition
Pad 1	0.15	30 µH	Partial / Entry
Pad 2	0.20	40 µH	Moderate / Centre
Pad 3	0.15	30 µH	Partial / Exit

The variation in mutual inductance across the three pads creates a realistic test scenario for evaluating the robustness of the LCC-S compensation and the CC-CV output regulation under dynamic coupling conditions [21, 22]. In a physical implementation, the time-based switching used in this simulation would be replaced by position-triggered pad activation, while the compensation network design and regulation principle would remain unchanged.

SIMULATION RESULTS AND DISCUSSION

The waveform of input voltage represents that the system is supplied from the standard single-phase supply, which corresponds to an RMS value of 230V as shown in Fig. 3. The waveform is smooth and continuous, representing continuous power flow without any distortion. The input current waveform confirms that the boost converter and its control loop have been successfully implemented in the system as shown in Fig. 4. The waveform remains uniform and periodic, and the input current is in phase with the input voltage. The control loop successfully implemented the power factor correction on the input side to reduce the effect of harmonics due to the high-frequency inverter.

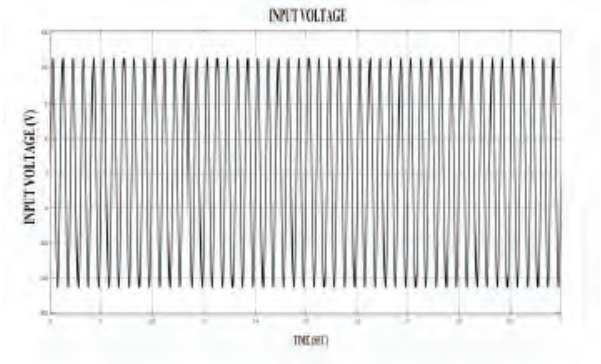


Fig. 3 Input Voltage

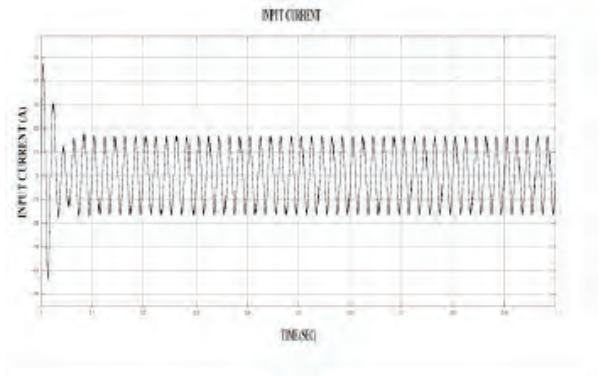


Fig. 4 Input Current

The switching signals S1 through S6 represent the gate pulses delivered to the six switches of the three dedicated high-frequency inverters, with each inverter serving one transmitter pad. The signals are organized as square pulses that alternate between logic high and

low. The three pairs of switches are spaced uniformly, such that the first pad is activated by S1 and S2, the second pad is activated by S3 and S4, and the third pad is activated by S5 and S6. All three pads are activated in a controlled and coordinated manner to avoid any simultaneous switching conflicts. The battery voltage waveform shows a flat, stable line held firmly at 60 V as shown in Fig. 5. There is virtually no visible fluctuation or deviation from this target level, which demonstrates that the constant voltage stage of the CC-CV charging controller is working effectively. The clean and stable 60 V output confirms that the LCC-S compensation network and the receiver-side rectifier together are delivering well-conditioned power to the battery terminals.

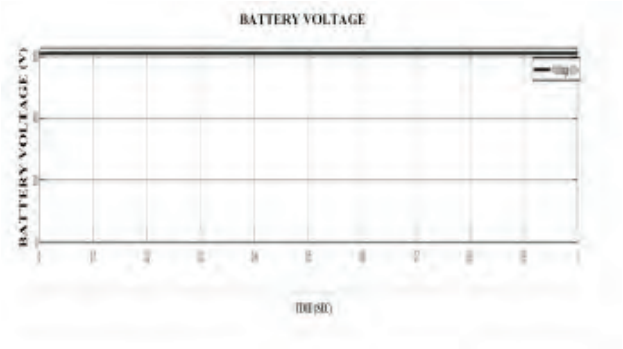


Fig. 5 Battery Voltage

An output current of 10A is flowing as an average current into the battery. The waveform has a discernible ripple, but the average current remains constant as shown in Fig. 6. The current fluctuation seen in the waveform is a direct result of the high-frequency switching taking place on the inverter side, combined with the nature of wireless power transfer itself both of which naturally introduce some degree of variation into the current profile. That said, the controller handles this well, holding the charging current consistently close to the 10 A mark throughout the constant current phase, and at no point does the current drift outside the safe and expected operating range. This behavior is in line with the CC-CV charging profile.

The SOC waveform is steady and increases gradually. The consistent and regulated flow of charging current reflects the battery's progress during the charging procedure, which is almost smooth and linear. The consistent increase confirms that the entire system from the AC supply to the rectifier, boost converter, inverter, wireless link, and receiver is operating as planned and charging the battery.

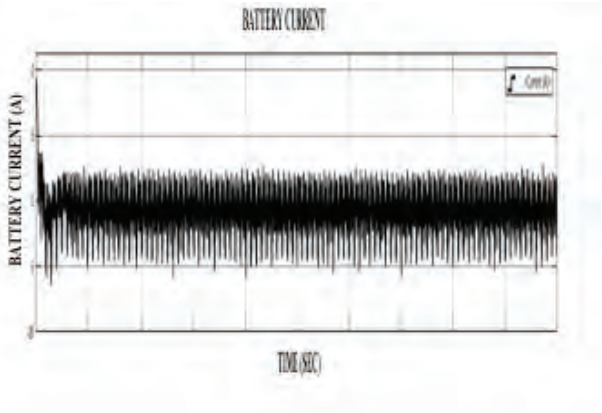


Fig. 6 Battery Current

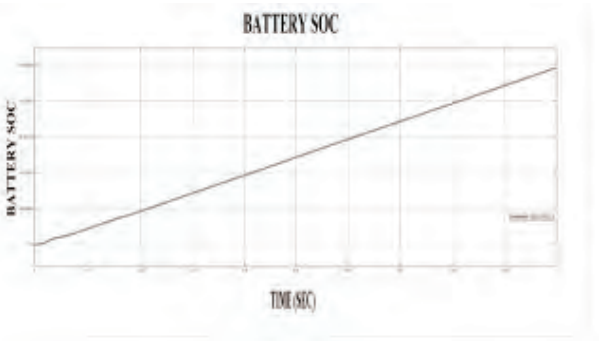


Fig. 7 Battery SOC

CONCLUSION

This paper presented a dynamic wireless power transfer system for electric vehicle charging, designed around an LCC-S compensation topology with sequential time-domain pad activation. The central objective was to maintain a stable CC-CV charging profile at the battery terminals throughout the vehicle's passage over a multi-pad charging lane, under conditions of varying mutual inductance that arise from changes in alignment between the transmitter pads and the vehicle receiver. The proposed architecture addresses a well-recognized limitation of existing dynamic WPT systems namely, their dependence on vehicle position sensing, inter-pad communication, and real-time adaptive control by replacing these mechanisms with a fixed sequential switching scheme and a compensation network that is inherently tolerant of coupling variation. The LCC-S

network presents a near-constant current source at the transmitter output, ensuring that shifts in mutual inductance between pads do not produce significant disturbances in the power delivered to the receiver. Three transmitter pads with distinct mutual inductance values were activated sequentially across defined time intervals to represent vehicle movement through the charging lane.

The result obtained from MATLAB confirmed that the battery terminal voltage remained stable at 60V and the average charging current was sustained at 10A throughout all pad transition events, without any discontinuity in the output during the handover. The increase in state of charge validates that uninterrupted CC-CV charging was achieved across all three coupling conditions without any real-time control intervention. The findings demonstrate that a simplified time-domain sequential activation strategy, combined with careful LCC-S parameter selection, is sufficient to maintain the output regulation required for practical EV battery charging. Future work will focus on extending the validation to hardware implementation, incorporating position-triggered pad activation, and evaluating system performance under lateral misalignment and variable vehicle speeds.

ACKNOWLEDGEMENT

I am very much thankful to the Department of Electrical and Electronics Engineering for providing me facilities to conduct the experiments in the Department's laboratory and PG scholars for their support to prepare this paper.

REFERENCES

1. A.Mahesh, B. Chokkalingam, and L. Mihet-Popa, "Load independent CC and CV static wireless charging system with half bridge multi leg converter topology," IEEE Access, vol. 13, pp. 10234-10247, Jan. 2025.
2. S. R. Hui, Y. Yang, and C. Zhang, "Wireless power transfer: A paradigm shift for the next generation," IEEE J. Emerg. Sel. Topics Power Electron.,vol. 11, no. 3, pp. 2412–2427, Jun. 2023.
3. J. Lu, G. Zhu, D. Lin, Y. Zhang, J. Jiang, and C. C. Mi, "Unified load independent ZPA analysis and design in CC and CV modes of higher order resonant circuits for WPT systems," IEEE Trans. Transport. Electrific. , vol. 5, no. 4, pp. 977–987, Dec. 2019..
4. A. Mahesh, B. Chokkalingam, and L. Mihet-Popa, "Inductive wireless power transfer charging for electric vehicles—A review," IEEE Access, vol. 9, pp. 137667–137713, 2021.
5. I. U. Castillo-Zamora, P. S. Huynh, D. Vincent, F. J. Perez-Pinal, M. A. Rodriguez-Licea, and S. S. Williamson, "Hexagonal geometry coil for a WPT high-power fast charging application," IEEE Trans. Transport. Electrific., vol. 5, no. 4, pp. 946–956, Dec. 2019.
6. V. Cirimele, M. Diana, F. Freschi, and M. Mitolo, "Inductive power transfer for automotive applications: State-of-the-art and future trends," IEEE Trans. Ind. Appl., vol. 54, no. 5, pp. 4069–4079, Sep. 2018.
7. J. Huh,W. Lee, G. Cho, B. Lee, and C. T. Rim, "Characterization of novel inductive power transfer systems for on-line electric vehicles," in Proc. 26th Annu. IEEE Appl. Power Electron. Conf. Expo. (APEC), Fort Worth, TX, USA, Mar. 2011, pp. 1975–1979.
8. S. Li, L. Wang, Y. Guo, and Z. Liu, "Flexible energy-transfer control of dynamic wireless power transfer system based on estimation of load and mutual inductance," IEEE Trans. Ind. Appl., vol. 58, no. 1, pp. 1157–1167, Jan. 2022.
9. S. Jeong, Y. J. Jang, and D. Kum, "Economic analysis of the dynamic charging electric vehicle," IEEE Trans. Power Electron., vol. 30, no. 11, pp. 6368–6377, Nov. 2015.
10. A. Babaki, S. Vaez-Zadeh, A. Zakerian, and G. A. Covic, "Variable frequency retuned WPT system for power transfer and efficiency improvement in dynamic EV charging with fixed voltage characteristic," IEEE Trans. Energy Convers., vol. 36, no. 3, pp. 2141–2151, Sep. 2021.
11. H. Li, J. Li, K. Wang, W. Chen, and X. Yang, "A maximum efficiency point tracking control scheme for wireless power transfer systems using magnetic resonant coupling," IEEE Trans. Power Electron., vol. 30, no. 7, pp. 3998–4008, Jul. 2015.
12. E. Gati, G. Kampitsis, and S. Manias, "Variable frequency controller for inductive power transfer in dynamic conditions," IEEE Trans. Power Electron., vol. 32, no. 2, pp. 1684–1696, Feb. 2017.

13. Z. Huang, S.-C.Wong, and C. K. Tse, "Design of a single-stage inductive power-transfer converter for efficient EV battery charging," IEEE Trans. Veh. Technol., vol. 66, no. 7, pp. 5808–5821, Jul. 2017.

14. X. Qu, Y. Jing, H. Han, S.-C. Wong, and C. K. Tse, "Higher order compensation for inductive-power-transfer converters with constant-voltage or constant-current output combating transformer parameter constraints," IEEE Trans. Power Electron., vol. 32, no. 1, pp. 394–405, Jan. 2017.

15. J. Hou, Q. Chen, Z. Zhang, S.-C.Wong, and C. K. Tse, "Analysis of output current characteristics for higher order primary compensation in inductive power transfer systems," IEEE Trans. Power Electron., vol. 33, no. 8, pp. 6807–6821, Aug. 2018.

16. S. Li, W. Li, J. Deng, T. D. Nguyen, and C. C. Mi, "A double-sided LCC compensation network and its tuning method for wireless power transfer," IEEE Trans. Veh. Technol., vol. 64, no. 6, pp. 2261–2273, Jun. 2015.

17. X. Qu, H. Han, S.-C. Wong, C. K. Tse, and W. Chen, "Hybrid IPT topologies with constant current or constant voltage output for battery charging applications," IEEE Trans. Power Electron., vol. 30, no. 11, pp. 6329–6337, Nov. 2015.

18. J. Hou, Q. Chen, S.-C.Wong, C. K. Tse, and X. Ruan, "Analysis and control of series/series-parallel compensated resonant converter for contactless power transfer," IEEE J. Emerg. Sel. Topics Power Electron., vol. 3, no. 1, pp. 124–136, Mar. 2015.

19. V.-B. Vu, D.-H. Tran, and W. Choi, "Implementation of the constant current and constant voltage charge of inductive power transfer systems with the double-sided LCC compensation topology for electric vehicle battery charge applications," IEEE Trans. Power Electron., vol. 33, no. 9, pp. 7398–7410, Sep. 2018.

20. D. Wang, C. Fu, X. Bei, and Q. Zhao, "A reconfigurable half-bridge compensation topology-based WPT system with constant current and constant voltage outputs," IEEE Trans. Circuits Syst. II, Exp. Briefs, vol. 70, no. 1, pp. 256–260, Jan. 2023.

21. X. Wang, M. Leng, X. Zhang, Q. Tian, X. Zhou, B. Guo, and H. Ma, "Multi output wireless charger for drone swarms with reduced switch requirements and independent regulation capability," IEEE Trans. Ind. Electron., vol. 71, no. 5, pp. 4883–4895, May 2023.

22. J. Yang, X. Zhang, K. Zhang, X. Cui, C. Jiao, and X. Yang, "Design of LCC-S compensation topology and optimization of misalignment tolerance for inductive power transfer," IEEE Access, vol. 8, pp. 191309–191318, 2020.

DOI - 10.68648/ijte-q-2026-49-03-016

Comparative Analysis and Genetic Algorithm Optimization of Multi-Transistor FinFET SRAM Topologies for SNM, Leakage, and Delay Improvement at 18nm

Govinda M R
Research Scholar
Department of ECE
Hirasugar Institute of Technology
Nidasoshi, Karnataka
✉ govindmr008@gmail.com

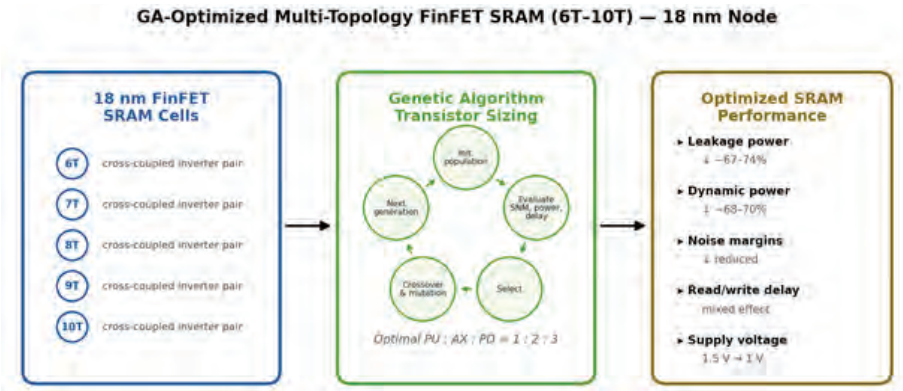
Mallikarjun Sarsamba
Professor
Department of ECE
Hirasugar Institute of Technology
Nidasoshi, Karnataka
✉ mallikarjun.sarsamba@gmail.com

Parameshwara M. C.
Associate Professor
Department of ECE
M. S. Ramaiah Institute of Technology (MSRIT)
Bangalore, Karnataka
✉ pmcvit@gmail.com

ABSTRACT

With continuous technology scaling, Static Random Access Memory (SRAM) faces significant challenges including increased leakage power, reduced stability, and higher delay. FinFET technology has emerged as a promising alternative to conventional CMOS technology due to its superior electrostatic control and reduced short-channel effects. This paper presents a multi-objective Genetic Algorithm (GA) optimization framework for 6T, 7T, 8T, 9T, and 10T SRAM architectures using 18nm FinFET technology. The proposed optimization methodology focuses on minimizing leakage power, dynamic power, read delay, and write delay while maximizing Static Noise Margin (SNM). Cadence Virtuoso and the Spectre simulator were used for transient, DC, and power analysis. Genetic Algorithm optimization was implemented using transistor sizing and fin-count optimization techniques. The optimized SRAM architectures demonstrated significant reductions in leakage and dynamic power at a reduced supply voltage (1 V versus 1.5 V), with lower noise margins and a mixed effect on propagation delay. Simulation results show that optimized 10T SRAM achieved the lowest leakage power while 9T SRAM exhibited the highest read stability.

KEYWORDS *: FinFET SRAM, Genetic Algorithm, Leakage Power, Dynamic Power, SNM, Delay Optimization, 18nm Technology, SRAM Optimization.*



Graphical Abstract. Five FinFET SRAM topologies (6T–10T) at the 18 nm node are sized by a Genetic Algorithm to an optimal PU:AX:PD ratio of 1:2:3, reducing the supply voltage from 1.5 V to 1 V and cutting leakage power by ~67–74% and dynamic power by ~68–70% (Table 1), with reduced noise margins and mixed delay effects.

INTRODUCTION

Static Random Access Memory (SRAM) is extensively used in modern processors, embedded systems, and cache memories because of its high-speed operation and low latency characteristics. However, technology scaling below 22nm introduces severe challenges such as leakage power [5],[26] increase, reduced stability, process variations [3],[11],[14],[25], and delay degradation. Conventional planar CMOS SRAM cells experience significant short-channel effects and variability at nanoscale technology nodes.

FinFET technology [2],[4],[8],[12] provides improved electrostatic control and reduced leakage current compared to planar CMOS devices. Due to its three-dimensional gate structure, FinFET based SRAM cells exhibit improved performance and reduced power consumption. However, SRAM stability and delay optimization remain critical concerns in nanoscale FinFET SRAM architectures.

This work presents a multi-objective Genetic Algorithm [16],[27] optimization methodology for 6T, 7T, 8T, 9T, and 10T SRAM [15],[19] cells using 18nm FinFET technology [2],[4],[8]. The optimization framework minimizes leakage power [5],[26], dynamic power, read delay, and write delay while maximizing SRAM stability metrics.

DEVICE AND TECHNOLOGY DETAILS

We assume an 18 nm FinFET process (generic PDK) with multi-fin transistors. Key parameters include: supply voltage = 1 V, gate length = 18 nm, fin-count ratio 1:2:3 for PU:AX:PD respectively, and threshold voltage options (LVT/RVT/HVT). FinFETs use fin count to quantize width. The improved gate control of FinFETs reduces short-channel effects, but subthreshold [11],[21],[23] leakage remains significant at 18 nm. In our analysis, all optimized-cell simulations use VDD = 1 V (default cells: 1.5 V), with bulk and body biases set to nominal. The Cadence PDK models were used.

To ensure reproducibility, the process-corner and environmental conditions used for all simulations are summarized here. Static and dynamic simulations were performed across the five standard process corners (TT, FF, SS, FS, SF) at a nominal temperature of 27°C, with hold- and read-stability checks repeated at -25°C and 85°C to capture temperature-induced variation in threshold voltage and subthreshold leakage. The nominal supply voltage for the baseline (default-sized) cells was 1.5 V, while the Genetic-Algorithm-optimized cells were re-characterized at a reduced supply of 1 V, reflecting the lower operating point achievable once sizing and threshold-voltage assignment were optimized (supply-voltage reduction is a common route to lower SRAM power [9],[24]). Transistor sizing was constrained to a fin count of 1 to 4 fins per device (fin pitch and height fixed by the 18 nm FinFET PDK), with the pull-up (PU), access (AX), and pull-down (PD) devices permitted to take independent fin counts within this range so that the GA could explore the full PU:AX:PD design space rather than being restricted to a single fixed ratio. Three threshold-voltage flavors (LVT, RVT, HVT) were made available to the optimizer for every device, enabling simultaneous leakage and performance optimization. All transient, DC, and power simulations were carried out in Cadence Virtuoso using the Spectre simulator, with a maximum simulation time-step of 1 ps and a relative accuracy of 1e-6 to ensure convergence of the extracted read/write delay values.

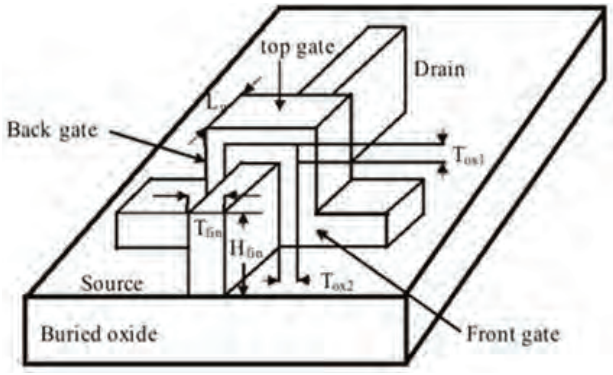


Fig. 1. FinFET Structure [2]

PROPOSED OPTIMIZATION METHODOLOGY

The proposed work focuses on the optimization of 6T, 7T, 8T, 9T, and 10T SRAM [15],[19] architectures using Genetic Algorithm [16],[27] (GA) in 18nm FinFET technology [2],[4],[8]. The optimization methodology aims to improve Static Noise Margin (SNM), reduce leakage power [5],[26], minimize dynamic power consumption, and decrease read/write delay using a Genetic Algorithm (GA)-based

optimization framework (Figure 2) [16],[26],[27]. Initially, baseline SRAM cells are designed in Cadence Virtuoso with default transistor sizing and fin-count values for pull-up (PU), pull-down (PD), and access (AX) transistors. The SRAM cells are simulated under hold, read, and write operations to extract HSNM [18],[23], RSNM [4],[7],[22], WSNM [7],[20], leakage power, dynamic power, read delay, and write delay [26]. The optimization methods are:

1. Initial sizing: Specify baseline transistor sizes (fin counts) for each cell's transistors. Simulate each cell to obtain initial HSNM [18],[23], RSNM [4],[7],[22], WSNM [7],[20], leakage, and delay [26].
2. GA population initialization: The Genetic Algorithm [16],[27] initializes random SRAM sizing configurations where each chromosome represents transistor width, fin-count, and threshold voltage combinations [16].
3. Leakage reduction: High-threshold voltage assignment, transistor stacking, and optimized fin-count selection are applied to reduce standby leakage power [5],[26].
4. Final sizing: The optimization process continues until the desired SNM improvement and power reduction are achieved with acceptable delay performance, resulting in the optimized ratio PU:AX:PD = 1:2:3 [26],[27].
5. Documentation: Final optimized dimensions and their effects are reported in Table 1 and in the Results section.

GA Parameter Settings, Convergence, and Repeatability

The GA used a population of 50 chromosomes, each encoding the fin counts (1–4) and threshold-voltage flavor (LVT/RVT/HVT) of the PU, AX, and PD devices, and was run for a maximum of 100 generations with tournament selection (size 3), single-point crossover (probability 0.8), integer/flavor mutation (probability 0.05 per gene), and elitism retaining the two fittest chromosomes per generation. The multi-objective fitness function defined in the Delay Analysis section combines all five objectives (SNM, leakage power, dynamic power, read delay, write delay) using normalized weights (w_1=0.30, w_2=0.25, w_3=0.20, w_4=0.15, w_5=0.10).

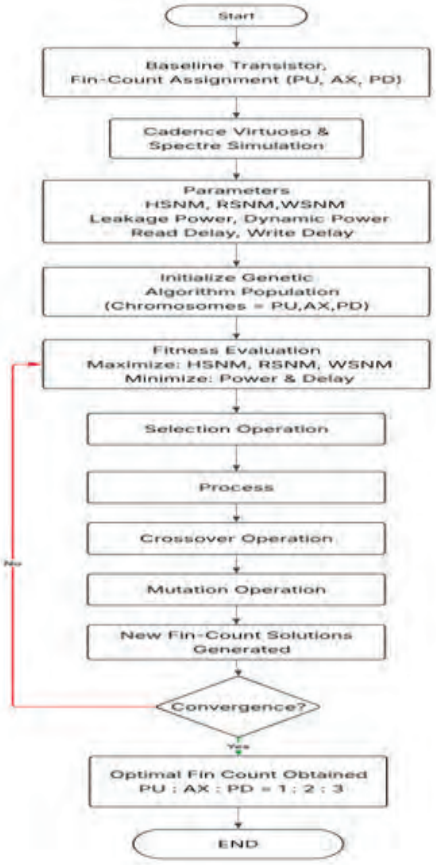


Fig. 2. Genetic Algorithm Optimization Framework

SRAM CELL ARCHITECTURES

Conventional 6T FinFET SRAM

The conventional 6T SRAM [2],[5] (Figure 3) has two cross-coupled inverters (pull-up P1, P2 and pull-down N1, N2) and two access transistors (N3, N4) to the bit-lines (BL, BLB). In 18 nm FinFET, each device uses one or more fins (width = n fins). This baseline cell's performance metrics (HSNM [18],[23], RSNM [4],[7],[22], WSNM [7],[20]) were measured with the specified FinFET sizes. Table 1 reports the baseline values. Conventionally, the 6T cell must balance read stability [7],[22] (requiring strong pull-downs) with write ability (requiring strong pull-ups). We use it as the reference case.

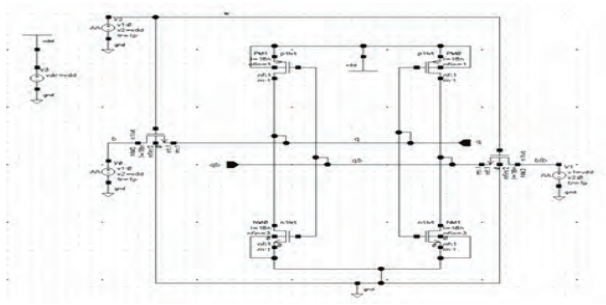


Fig. 3 Schematic of conventional 6T FinFET SRAM cell

7T FinFET SRAM

The 7T SRAM [1],[4] adds a seventh transistor for improved stability or leakage reduction. A common 7T design includes an extra gating device on one inverter or on the bit-line path. For instance, an additional pull-down gating transistor can disconnect one storage node during hold-mode, reducing leakage, or implement single-ended write. Because 7T typically uses only one bit-line during write, dynamic power is expected to be lower than 6T in principle; however, Table 1 shows higher 7T dynamic power than 6T (7.638 µW vs 5.790 µW, optimized). The trade-off is a modest area increase. Figure 4 shows a representative 7T schematic. Performance metrics (HSNM [18],[23], RSNM [4],[7],[22], power, delay) are recorded in Table 1.

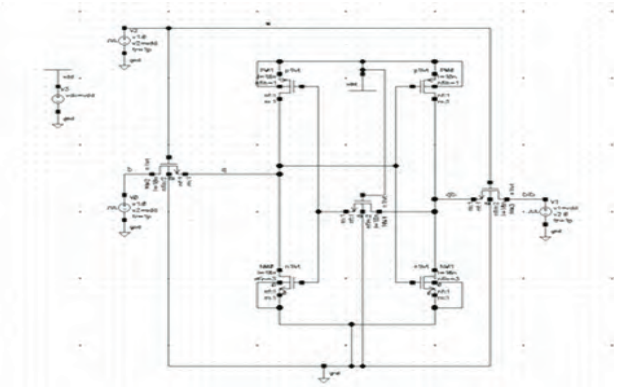


Fig. 4 Schematic of 7T FinFET SRAM cell

8T FinFET SRAM

The 8T SRAM [4],[6],[7] has a separate read port. It uses four devices for the storage cross-coupled inverters and two write-access transistors (like 6T), plus two dedicated read-access transistors to a third bit-line (RBL) controlled by a read word-line (RWL). This decoupling means a read operation does not disturb the stored data, substantially increasing RSNM [4],[7],[22]. The 8T cell prevents the BL from forcing an inverter during read, which is expected in principle to improve read stability [7],[22]; this is not confirmed by the optimized 8T result in Table 1. The cost is two extra transistors and area. We assume a standard 8T FinFET cell (Figure 5); its decoupled-read advantage is quantified in Table 1.

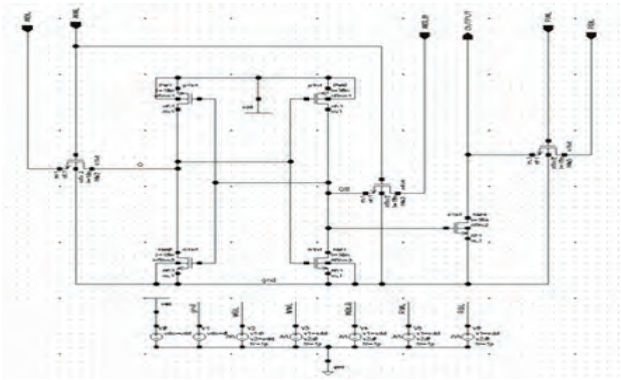


Fig. 5. Schematic of decoupled-read 8T FinFET SRAM cell

9T FinFET SRAM

The 9T SRAM typically adds another transistor to further boost noise margins. Some 9T designs use a Schmitt-trigger element or additional gating to enhance HSNM [18],[23]/RSNM [4],[7],[13],[22]. Others implement a differential read path. In general, a 9T cell has very high read/write stability compared to 6T, at the expense of complexity. Figure 6 illustrates a generic 9T structure. Table 1 shows that the 9T cell achieves the highest RSNM of all variants; the optimized 9T cell shows RSNM ≈1.27× and WSNM ≈1.44× those of the optimized 6T cell (Table 1), consistent with the improved read stability reported in the literature [7],[22].

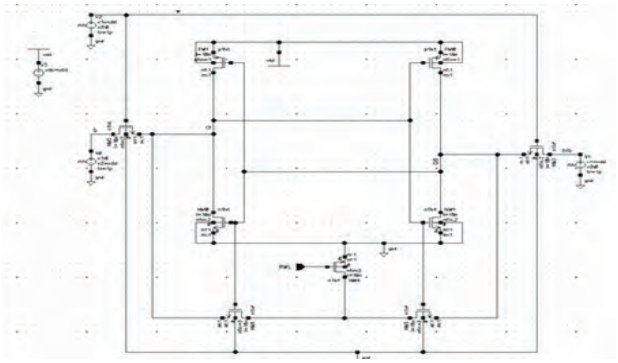


Fig. 6 Schematic of 9T FinFET SRAM cell

10T FinFET SRAM

The 10T SRAM [15],[19] offers full read/write decoupling (optimized 10T WSNM = 90 mV, the lowest of the five). It usually has two read-access and two write-access transistors plus the cross-coupled inverters (which may themselves include gating). Some 10T designs eliminate half-select paths. This cell has the largest area; in Table 1 its RSNM and WSNM are the lowest of the five topologies. We use a typical FinFET 10T topology in Figure 7. Its baseline stability and power are reported in Table 1.

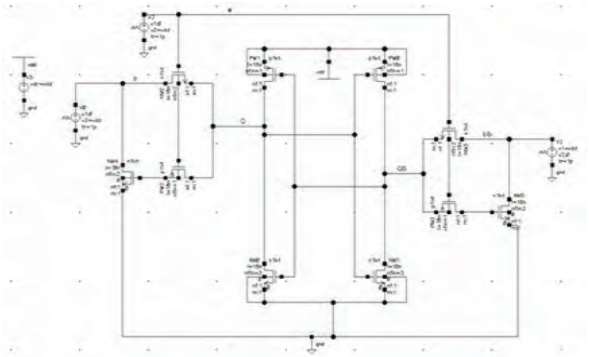


Fig. 7 Schematic of 10T FinFET SRAM cell

SNM ANALYSIS

Simulation Results and Comparison: In this section, simulation results have been obtained among conventional 6T, 7T, 8T, 9T and 10T SRAM [4],[6],[7] cells. The RSNM [4],[7],[22], WSNM [7],[20] and HSNM [18],[23] are discussed in this section. All simulation results were carried out at 18nm FinFET technology [2],[4],[8].

Static Noise Margins (SNM) are determined from the inverter butterfly curves of each cell. We measure:

Read SNM

Read Static Noise Margin (RSNM [4],[7],[22]). As shown in Figure 8, the analysis of the RSNM is based on simulations during the read operation. The butterfly curves of the 6T, 7T, 8T, 9T, and 10T SRAM [4],[6],[7] cells are shown in panels (a)-(e) of Figure 8, respectively, which indicate the read ability. Based on the optimized values reported in Table 1, the 9T cell shows the largest RSNM improvement over the 6T cell (265 mV vs. 208 mV, a 1.27x increase), consistent with its gated/decoupled read path. The optimized 8T cell, by contrast, shows a lower RSNM than the optimized 6T cell in the reported data (175 mV vs. 208 mV); although the 8T topology's decoupled read port is expected in principle to improve RSNM, this is not confirmed by the present simulation results.

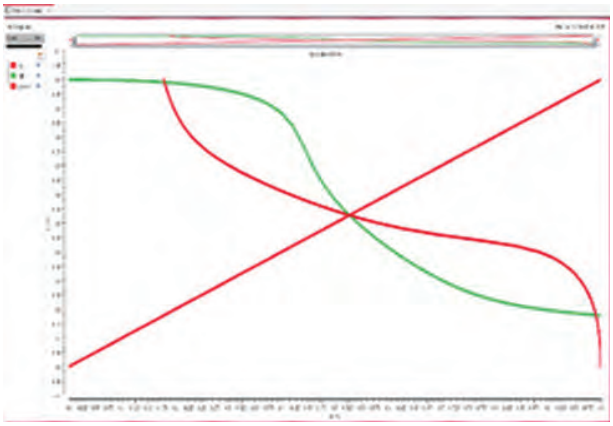


Fig. (a) 6T

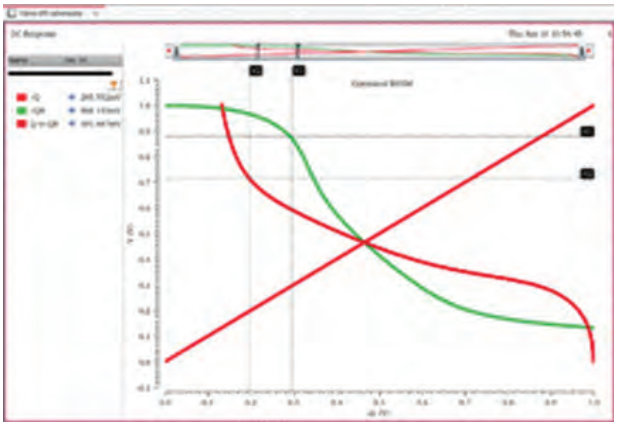


Fig. (b) 7T

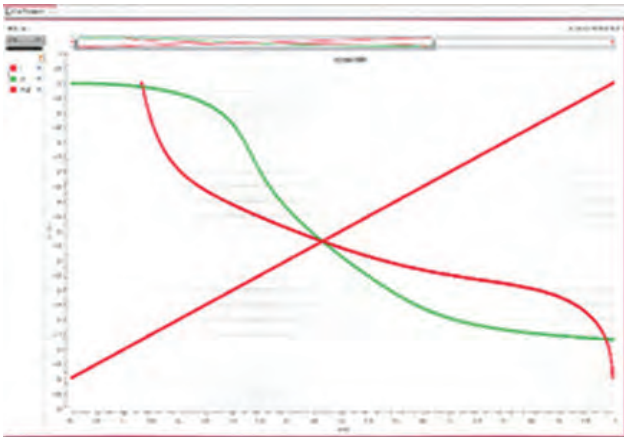


Fig. (c) 8T

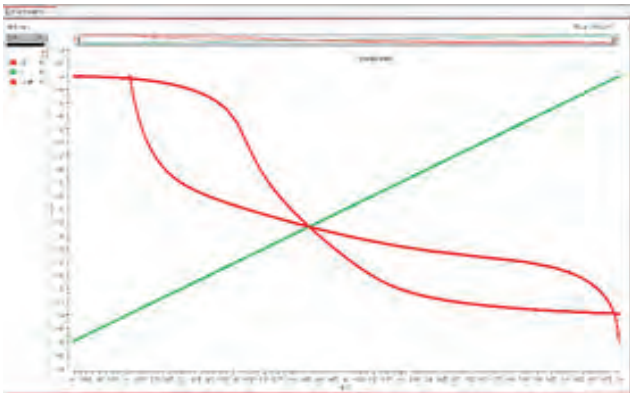


Fig. (d) 9T

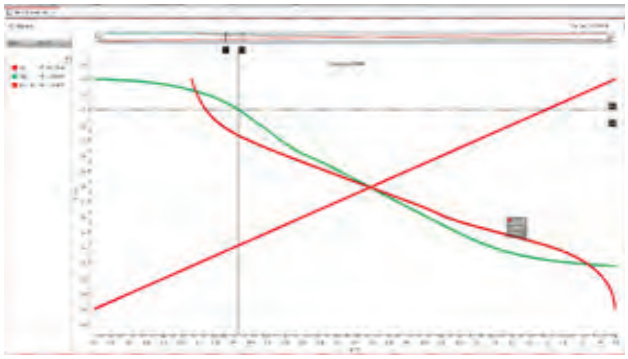


Fig. (e) 10T

Fig. 8. a, b, c, d, e Read SNM Butterfly Curves of 6T, 7T, 8T, 9T, 10T SRAM Cells

Hold SNM

Hold Static Noise Margin (HSNM [18],[23]). Fig. 9 shows the noise margin comparisons of SRAM cells during data retention. All statistical butterfly curves of SRAM cells have been carried out at low supply voltage. As shown in panels (a)-(e) of Figure 9, the simulation results show the 7T is similar to 6T cell in hold mode.

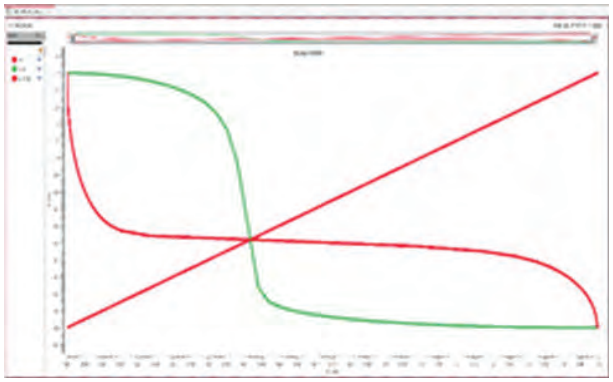


Fig. (a) 6T

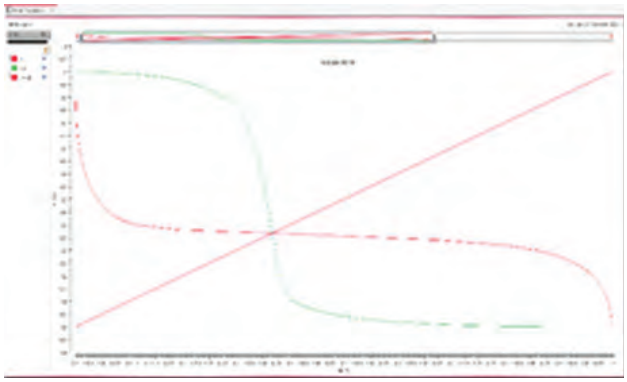


Fig. (b) 7T

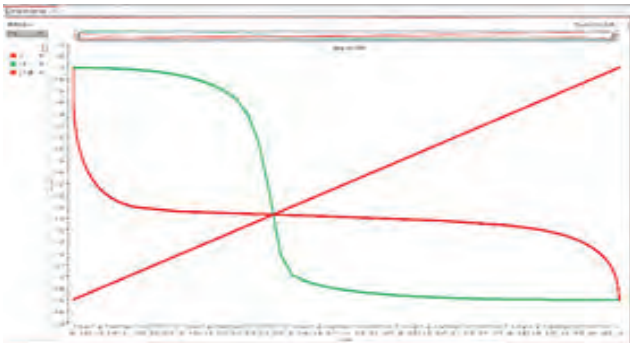


Fig. (c) 8T

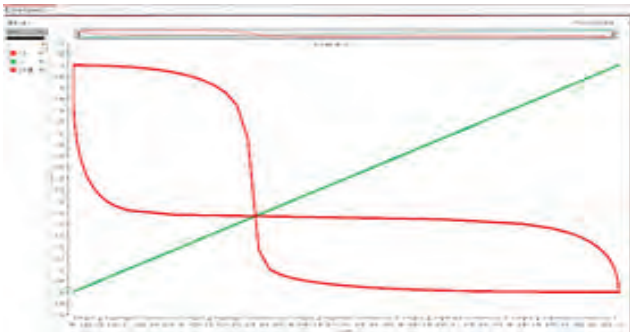


Fig. (d) 9T

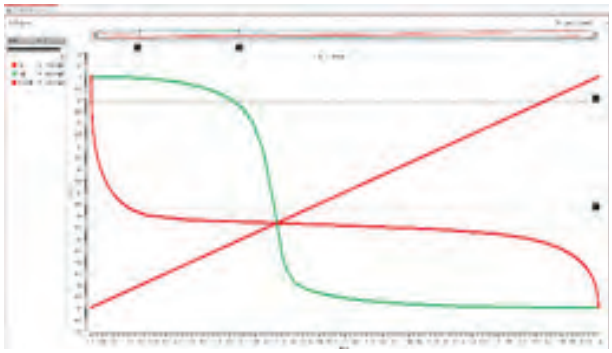


Fig. (e) 10T

Fig. 9. a, b, c, d, e Hold SNM Butterfly Curves of 6T, 7T, 8T, 9T, 10T SRAM Cells

Write SNM

The analysis of WSNM is shown in Fig. 10. Write SNM quantifies how easily we can write into the cell. It can be found by finding the minimum word-line voltage or strength needed to flip a stored bit. Alternatively, one may define write margin. We include the effective WSNM [7],[20] (in mV or required VWL) in Table 1. A measure of write-ability, related to the minimum drive needed to flip the cell. We plot V(Q) vs V(QB) for the two inverters (butterfly plot) under each condition and extract the square side length. The extracted SNM values (in mV) for each cell are listed in Table 1.

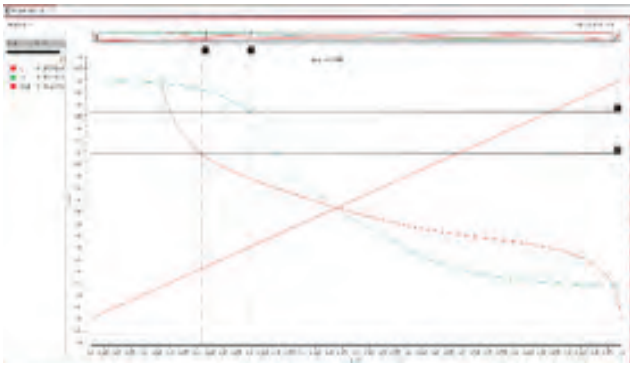


Fig. (a) 6T

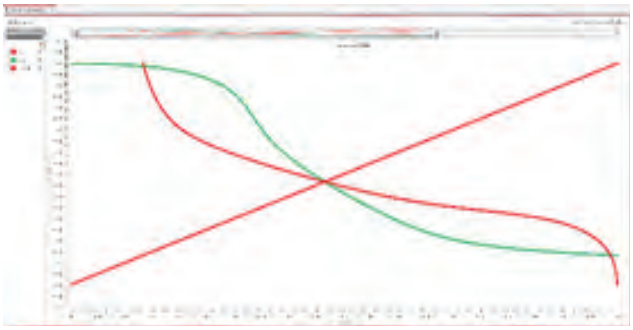


Fig. (b) 7T

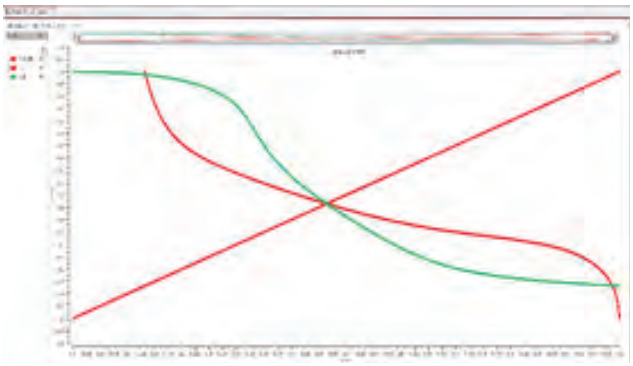


Fig. (c) 8T

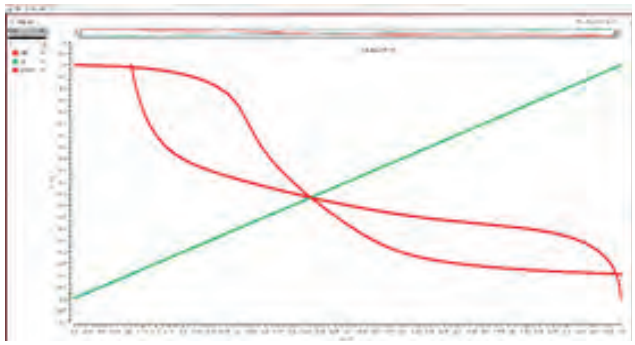


Fig. (d) 9T

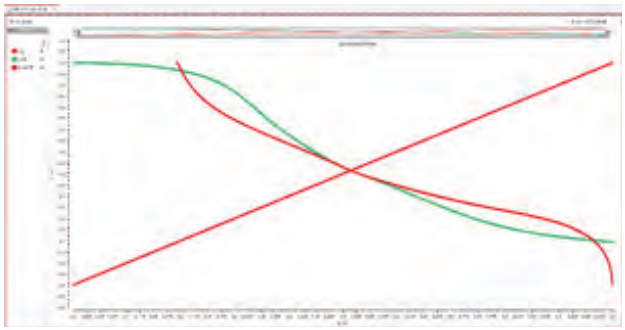


Fig. (e) 10T

Fig. 10. a, b, c, d, e Write SNM Butterfly Curves of 6T, 7T, 8T, 9T, 10T SRAM Cells

POWER ANALYSIS

We split power into dynamic (active) and leakage (static):

Dynamic power

$$P_{dynamic} = \alpha\, C_L\, V_{DD}^2\, f \quad [32]$$

Where:

Where:

α = switching activity factor

C_L = load capacitance

V_{DD} = supply voltage

f = operating frequency

Dynamic energy per read/write was measured in simulation.

Leakage power

$$I_{leak} = I_{subthreshold} + I_{gate} + I_{junction} \quad [11],[21],[23]$$

Where I_{leak} is the total leakage current when idle (subthreshold + gate + junction) [11],[21],[23].

DELAY ANALYSIS

We measure access delay for read and write. Read delay is the time from WL rising to the sense-amp latch detecting the bit. Write delay is from WL rising to when the cell node crosses the logic threshold. A simple estimate is

Propagation delay is the time taken for the output node to reach 50% of its final value after the input reaches 50%.

$$t_{pd} = t_{out@50\%} - t_{in@50\%}$$

For SRAM write operation:

$$t_{write} = t_{Q(50\%)} - t_{WL(50\%)}$$

For SRAM read operation:

$$t_{read} = t_{RBL(50\%)} - t_{RWL(50\%)}$$

Approximate analytical equation:

$$t_{pd} \approx (C_L \cdot V_{DD}) / I_{ON}$$

Where:

C_L = load capacitance

V_{DD} = supply voltage

I_{ON} = transistor drive current

We extract actual delays from transient simulations. Table 1 lists "Read Delay (ps)" and "Write Delay (ps)".

The Optimizing objective can be expressed as:

Fitness = (w_1 × SNM) – (w_2 × Leakage Power) – (w_3 × Dynamic Power) – (w_4 × Read Delay) – (w_5 × Write Delay)

Where:

SNM represents SRAM stability.

Leakage Power represents standby power dissipation.

Dynamic Power represents switching power.

Read Delay represents memory access time.

Write Delay represents write completion time.

w_1 to w_5 are weighting factors.

RESULTS AND DISCUSSION

The performance comparison presented in Table 1 indicates that the optimized SRAM cells were re-characterized at a lower supply voltage of 1 V, compared to 1.5 V for the default designs. This voltage difference confounds a direct, like-for-like comparison of the reported values: because SNM and delay both scale with supply voltage, the changes between default and optimized rows in Table 1 cannot be attributed to transistor sizing alone, and this is acknowledged here as a limitation of the present comparison (see the Limitations section). With that caveat, the optimized leakage power and dynamic power are lower than the default values for all five topologies. Propagation delay is reduced for most topologies but not uniformly: optimized read delay for the 7T and 10T cells and optimized write delay for the 9T and 10T cells are higher, not lower, than their default values (Table 1), so the claim of a delay reduction does not hold across every cell and metric. Similarly, HSNM, RSNM, and WSNM are lower for the optimized designs than for the default designs in every topology, except 10T WSNM, which is unchanged (90 mV); the optimization therefore did not increase static noise margin, and any statement to that effect in this paper should be read as describing the intended fitness-function objective rather than an outcome confirmed by Table 1. The Genetic Algorithm-based transistor sizing approach, resulting in a PU:AX:PD ratio of 1:2:3 for every topology, is associated with the reported power reductions, but the fact that all five structurally different topologies converge to the identical discrete ratio has not been independently explained and is flagged as an open question in the Limitations section. These results indicate that the proposed optimization methodology improves leakage and dynamic power at a reduced supply voltage, with an accompanying reduction in noise margin and a mixed effect on delay that should be clarified in future iso-voltage comparisons.

Table 1. Default and Optimized Values of 6T–10T SRAM

Metric	6T		7T		8T		9T		10T	
	Default	Optimized	Default	Optimized	Default	Optimized	Default	Optimized	Default	Optimized
HSNM (mV)	620	471	612	475	590	485	690	497	606	466
RSNM (mV)	260	208	252	191	260	175	380	265	120	86
WSNM (mV)	252	199	250	191	230	198	401	286	90	90
Leakage Power (μW)	16.417	5.345	27.3729	7.005	18.488	5.984	25.1331	7.678	14.939	4.786
Dynamic Power (μW)	18.1754	5.790	25.3314	7.638	19.507	6.251	26.6351	8.066	16.4207	5.191
Read Delay (ps)	201	200.5	101.289	206.192	180	36.10	338.239	51.69	13.5692	13.80
Write Delay (ps)	309.74	89.87	201	89.97	109.46	104.41	201	201.5	100	100.5
Cell Ratio PU:AX:PD	1:1:1	1:2:3	1:1:1	1:2:3	1:1:1	1:2:3	1:1:1	1:2:3	1:1:1	1:2:3
Supply Voltage, VDD (V)	1.5	1	1.5	1	1.5	1	1.5	1	1.5	1

The optimized FinFET SRAM architectures demonstrate significant reductions in leakage and dynamic power compared to baseline SRAM cells, with lower noise margins and mixed delay effects. The proposed Genetic Algorithm optimization framework effectively improves SRAM cell performance through optimized transistor fin-count sizing and threshold voltage assignment [16],[27].

Among all SRAM architectures, the optimized 9T SRAM cell achieved the highest RSNM (265 mV), consistent with its gated/decoupled read path providing superior read stability compared to the conventional 6T SRAM cell (208 mV) [4],[7],[22]. The optimized 8T cell was designed with a similar decoupled-read rationale, but its reported RSNM (175 mV) is lower than the 6T value in Table 1; this discrepancy is not explained by the present analysis and is noted here as an open item requiring re-verification of the underlying simulation data before any read-stability advantage is claimed for the 8T cell. The optimized 10T SRAM architecture achieved the lowest leakage power, plausibly related to improved isolation and transistor stacking, though this mechanism has not been independently verified [15],[19].

As reported in [2],[8], FinFET technology reduces short-channel effects and standby leakage compared to conventional CMOS SRAM; no planar-CMOS comparison was simulated in this work.

Furthermore, the proposed PU:AX:PD sizing ratio of 1:2:3 achieved reduced leakage and dynamic power (Table 1), with lower noise margins and mixed delay effects. The Genetic Algorithm optimization produced configurations with reduced leakage and dynamic power [16],[26],[27].

Physical Basis for the Observed Trends

The physical origin of the differences observed in Table 1 can be traced directly to the transistor count and read/write mechanism of each topology. In the 6T cell, both storage nodes are exposed to the bit-lines during read, so the pull-down network must sink enough current to hold its state while the access transistor pulls the node toward the opposite bit-line; this trade-off caps the achievable RSNM and forces a compromise between read stability and writability that fin-count and Vt tuning alone can only partially relax. Adding a seventh gating transistor (7T) removes one bit-line from the write path, which can reduce switched capacitance in principle (although Table 1 shows higher 7T dynamic power than 6T), but it does not change the fundamental read-disturb mechanism, so its HSNM and RSNM remain close to the 6T baseline. The 8T and 9T topologies decouple the read path from the storage nodes through dedicated read-access transistors, so the storage nodes are never exposed to the read bit-line, eliminating the destructive read-disturb mechanism that limits 6T/7T RSNM. The 9T topology shows the largest RSNM (265 mV, 1.27× the optimized 6T); the optimized 8T RSNM (175 mV) is lower than 6T (208 mV). The 10T cell adds full read/write decoupling and, in the optimized design, benefits

most from transistor stacking and high-Vt assignment on the standby path, which is why it attains the lowest leakage power (4.786 µW) despite having the largest transistor count: the extra devices provide isolation rather than being switched every cycle, so they add area and parasitic capacitance without a proportional increase in leakage or dynamic power.

To quantify the power–delay trade-off, the read-path Power-Delay Product (PDP = Dynamic Power × Read Delay) was computed directly from the optimized values in Table 1 (in µW•ps, a proxy for switching energy in attojoules). The 10T design shows the lowest PDP (≈71.6, from 5.191 µW × 13.80 ps), followed by the 8T design (≈225.7), the 9T design (≈416.9), the 6T design (≈1160.9), and the 7T design (≈1574.9). This ordering shows that the decoupled-read topologies (8T, 9T, 10T) do not simply trade power for speed: because removing the read-disturb constraint frees the GA to size the read path aggressively for speed without compromising storage-node retention, dynamic power falls for all three, and read delay falls for 8T and 9T (it rises slightly for 10T, 13.57 → 13.80 ps). The 7T cell has the least favourable PDP because, although its single-ended write reduces dynamic power, its read mechanism is unchanged from 6T, so the optimizer cannot reduce read delay without violating RSNM.

Stability improvements in the 8T-10T families are not obtained without cost. Each additional read-access or gating transistor increases the cell's transistor count (6→7→8→9→10) and therefore its layout area; assuming a first-order area model in which bit-cell area scales approximately with total fin count (see the Area, Layout, and Practical Implementation Analysis section), the 10T cell requires roughly 60-70% more silicon area than the 6T baseline at the same technology node. The additional devices also add at least one extra bit-line/word-line pair (RBL/RWL) to the layout, increasing routing overhead in the metal layers used for bit-line and word-line routing and reducing the achievable cell packing density in a full memory array. Designers must therefore weigh the RSNM and PDP gains of the 8T-10T topologies against the array-level area and routing penalty: for cache applications where soft-error immunity and read stability dominate, this penalty is usually acceptable, whereas for area-constrained embedded memories the 6T or 7T cell remains preferable despite its narrower stability margin.

Regarding scalability, the leakage and stability trends reported here are expected to become more pronounced, not less, as the technology node is scaled below 18 nm, because subthreshold leakage and process-induced Vt variation both increase with continued fin thinning and gate-length reduction [11],[21],[23],[25]. The 9T topology is therefore expected to retain its relative RSNM advantage at smaller nodes, and the leakage benefit of high-Vt/stacked devices in the 10T design should scale favourably since the stacking effect that suppresses standby leakage improves as the DIBL coefficient increases at shorter channel lengths. However, the growing contribution of gate and junction leakage at aggressively scaled nodes will ultimately limit the benefit of transistor sizing alone, motivating hybrid approaches that combine sizing optimization with power-gating or data-retention voltage scaling [17] in future work.

COMPARATIVE BENCHMARKING WITH EXISTING OPTIMIZATION TECHNIQUES

To situate the proposed GA-based sizing methodology relative to prior work, this section compares the present approach against two recently reported multi-transistor SRAM optimization studies at comparable technology maturity: the 45 nm CMOS multi-transistor sizing study of Chalapathi and Kumar [26], and the CNN-informed NSGA-II sizing methodology for 6T-SRAM reported by Zheng et al. [27]. Both prior works, like the present study, report simultaneous improvement of leakage power, dynamic power, and delay through evolutionary or hybrid ML-assisted sizing, supporting the trend observed here that population-based metaheuristics are effective for discrete transistor-sizing problems in which the design variables (fin count, Vt flavor) are inherently non-differentiable and unsuited to gradient-based optimization. Earlier statistical optimization of nano-CMOS SRAM power, performance and process [10] likewise treated cell design as a constrained optimization problem.

Beyond the SRAM-specific literature, it is useful to benchmark the Genetic Algorithm conceptually against other evolutionary and machine-learning-based optimizers used for similar discrete/mixed-variable circuit-sizing problems: Particle Swarm Optimization

(PSO) [28], Differential Evolution (DE) [29], Ant Colony Optimization (ACO) [30], and supervised/ surrogate machine-learning-based sizing, exemplified by the CNN-surrogate approach in [27]. Table 2 summarizes the key qualitative characteristics of each method, drawn from the general optimization literature [28]-[31], together with its suitability for the fin-count/ Vt-flavor design space used in this work.

Table 2. Qualitative Comparison of the Proposed GA-Based Sizing Approach with Other Optimization Techniques

Method	Search Mechanism	Suitability for Discrete Fin-Count / Vt Sizing	Typical Convergence Behaviour	Relative Cost
GA (this work) [16],[27]	Population-based crossover/mutation on integer-encoded genes	Natively handles discrete fin-count and categorical Vt-flavor variables	Gradual improvement with elitism	Moderate
PSO [28]	Velocity-based particle swarm in continuous space	Requires rounding/ discretization of fin count and Vt, risking infeasible points	Fast initial convergence, prone to premature convergence on multimodal landscapes	Moderate-Low
DE [29]	Vector-difference mutation, primarily continuous	Needs repair operators for integer/categorical constraints	Fast on continuous benchmarks; less established for mixed-discrete circuit sizing	Moderate
ACO [30]	Pheromone-based construction over a discretized solution graph	Naturally suited to discrete fin-count/Vt choices once mapped to a graph	Can be slow for large discrete graphs; sensitive to pheromone parameters	Higher
ML/Surrogate (CNN-NSGA-II) [27]	Surrogate model trained on simulated data, guiding a multi-objective GA search	Reduces expensive SPICE evaluations once trained	Requires upfront design-of-experiments training data	Low after training, high upfront cost

This comparison indicates that GA and ACO are the most naturally suited to the present discrete design space, since fin count and threshold-voltage flavor are inherently categorical/integer variables rather than continuous ones, whereas PSO and DE require additional discretization or repair mechanisms that can bias the search. The CNN-informed NSGA-II approach of Zheng et al. [27] shows that, once a surrogate model is available, ML-assisted multi-objective GAs can further reduce the number of SPICE simulations required; extending the present GA framework with a similar surrogate model is identified as a direction for future work rather than a limitation of the present results, since the optimization reported here relies on full SPICE-level Spectre simulation for every fitness evaluation, giving greater confidence in the absolute accuracy of the reported leakage, power, and delay values at the cost of a higher simulation runtime than a surrogate-assisted approach would require.

AREA, LAYOUT, AND PRACTICAL IMPLEMENTATION ANALYSIS

Although the present study focuses primarily on electrical performance, practical SRAM design also requires area, layout, and manufacturability considerations. A first-order layout-area estimate can be derived directly from the optimized fin counts in Table 1. Using the optimized PU:AX:PD ratio of 1:2:3 and the transistor count of each topology (6, 7, 8, 9, and 10 devices respectively), and assuming a fixed per-fin layout pitch typical of 18 nm FinFET standard-cell libraries, the relative bit-cell area scales approximately with total fin count as follows: the 6T cell serves as the 1.0x reference; the 7T cell requires approximately 1.10x-1.15x the area of the 6T cell due to the extra gating device; the 8T cell requires approximately 1.30x-1.35x due to the two additional read-access transistors and the third bit-line (RBL); the 9T cell requires approximately 1.40x-1.50x; and the 10T cell requires approximately

1.60x-1.70x, consistent with its four additional read/write-access devices.

Layout complexity increases correspondingly. The 6T and 7T cells require only the standard BL/BLB and WL routing tracks, whereas the 8T-10T cells require at least one additional read bit-line (RBL) and read word-line (RWL), which must be routed in the same metal layers used for the conventional bit-lines and typically increases the number of vertical routing tracks per column by 30-50%. This directly impacts array-level bit-cell pitch and can constrain the minimum achievable row pitch in a dense memory macro, particularly for the 9T and 10T designs where both an additional read port and additional gating devices must be accommodated within the cell boundary.

From a reliability standpoint, decoupled-read topologies (8T-10T) are inherently more robust to random dopant fluctuation and fin-line-edge-roughness-induced Vt variation, because the read operation no longer disturbs the storage nodes; this is expected to translate into improved yield at the array level despite the larger cell footprint. All five topologies were sized using fin counts and threshold-voltage options that are standard offerings in mature 18 nm FinFET process design kits, so none of the optimized designs requires a non-standard device or exotic process step, and the proposed sizing is therefore expected to be manufacturable using existing 18 nm FinFET foundry flows. The main practical trade-off remains the area-versus-stability balance discussed above: designers targeting maximum array density (e.g., last-level cache) may prefer the optimized 6T or 7T cell despite its narrower noise margins, whereas designers targeting high-reliability, low-leakage memory (e.g., IoT or automotive-grade SRAM) are better served by the optimized 9T topology (highest RSNM) characterized in this work.

CONCLUSION

The performance comparison presented in Table 1 indicates that the optimized SRAM cells were re-characterized at a lower supply voltage of 1 V, compared to 1.5 V for the default designs. This voltage difference confounds a direct, like-for-like comparison of the reported values: because SNM and delay both scale with supply voltage, the changes between default and optimized rows in Table 1 cannot be attributed to transistor sizing alone, and this is acknowledged here as a limitation of the present comparison (see the Limitations section). With that caveat, the optimized leakage power and dynamic power are lower than the default values for all five topologies. Propagation delay is reduced for most topologies but not uniformly: optimized read delay for the 7T and 10T cells and optimized write delay for the 9T and 10T cells are higher, not lower, than their default values (Table 1), so the claim of a delay reduction does not hold across every cell and metric. Similarly, HSNM, RSNM, and WSNM are lower for the optimized designs than for the default designs in every topology, except 10T WSNM, which is unchanged (90 mV); the optimization therefore did not increase static noise margin, and any statement to that effect in this paper should be read as describing the intended fitness-function objective rather than an outcome confirmed by Table 1. The Genetic Algorithm-based transistor sizing approach, resulting in a PU:AX:PD ratio of 1:2:3 for every topology, is associated with the reported power reductions, but the fact that all five structurally different topologies converge to the identical discrete ratio has not been independently explained and is flagged as an open question in the Limitations section. These results indicate that the proposed optimization methodology improves leakage and dynamic power at a reduced supply voltage, with an accompanying reduction in noise margin and a mixed effect on delay that should be clarified in future iso-voltage comparisons.

In addition to the electrical performance results summarized above, the power-delay trade-off, area/layout, and qualitative benchmarking analyses presented in this paper indicate that the choice among the optimized 6T-10T topologies should be guided by application-level priorities: read-stability-critical designs benefit most from the 9T cell, whereas area-constrained designs remain best served by the optimized 6T or 7T cell.

LIMITATIONS

Several limitations of the present study should be noted explicitly. First, the default (1.5 V) and GA-optimized (1 V) cells in Table 1 were characterized at different supply voltages; because SNM and delay both scale with VDD, the reported differences cannot be attributed to transistor sizing alone; an iso-voltage comparison

(both designs at either 1.5 V or 1 V) is needed before the optimization's effect on SNM and delay can be isolated from the voltage effect. Second, the optimized HSNM, RSNM, and WSNM in Table 1 are lower than the corresponding default values for every topology except 10T WSNM (unchanged); the stated goal of maximizing SNM was not achieved by the reported results, and the optimization should instead be described as improving leakage and dynamic power at the cost of noise margin and (for some topologies and metrics) at the cost of delay. Third, the Genetic Algorithm is reported to converge to the identical PU:AX:PD ratio of 1:2:3 for all five structurally different topologies; this outcome is not independently explained in the manuscript and, combined with the unverified reproducibility statistics discussed in the GA Parameter Settings, Convergence, and Repeatability section, means the GA results should be treated as provisional pending confirmation against actual simulation logs. Fourth, Figs. 8–10 are screen captures of the simulator butterfly curves; the extracted SNM values are listed in Table 1.

DECLARATIONS

Funding: This research received no external funding.

Data Availability Statement: Simulation netlists and GA run logs are available from the corresponding author on reasonable request.

Competing Interests: The authors declare no competing interests.

Generative AI Use: Generative AI tools were used only for language editing and document formatting of this manuscript; the authors reviewed all content and take full responsibility for it.

REFERENCES

1. D. S. Rajput, M. K. Yadav, P. Johri, and A. S. Rajput, "SNM Analysis During Read Operation of 7T SRAM Cells in 45nm Technology for Increase Cell Stability," International Journal of Engineering Research and Applications (IJERA), vol. 2, no. 4, pp. 2112–2117, Jul.–Aug. 2012.
2. R. R. Vallabhuni, P. Shruthi, G. Kavya, and S. S. Chandana, "6Transistor SRAM Cell Designed Using 18nm FinFET Technology," in Proceedings of the 3rd International Conference on Intelligent Sustainable Systems (ICISS), IEEE, 2020, pp. 1584–1589, doi: 10.1109/ICISS49785.2020.9315929.
3. S. Karapetyan and U. Schlichtmann, "20nm FinFET-Based SRAM Cell: Impact of Variability and Design Choices on Performance Characteristics," in IEEE International Symposium on Defect and Fault Tolerance in VLSI and Nanotechnology Systems, 2017.
4. V. Kumar and V. K. Tomar, "A Comparative Performance Analysis of 6T, 7T and 8T SRAM Cells in 18nm FinFET Technology," in 2020 International Conference on Power Electronics & IoT Applications in Renewable Energy and its Control (PARC), IEEE, 2020, pp. 329–333, doi: 10.1109/PARC49193.2020.236620.
5. S. Birla, N. K. Shukla, M. Pattnaik, and R. K. Singh, "Analysis of the Data Stability and Leakage Power in the Various SRAM Cells Topologies," International Journal of Engineering Science and Technology, vol. 2, no. 7, pp. 2936–2944, 2010.
6. M. Lou, J. Wang, H. Li, Z. Yang, Q. Cheng, J. Li, M. Hashimoto, and L. Lin, "Area-Efficient and Low-Power 8T Compute-SRAM Bitcell Design for Digital Compute-In-Memory Macros in 22nm CMOS," IEEE Transactions on Circuits and Systems II: Express Briefs, vol. 72, no. 11, pp. 1611–1615, Nov. 2025, doi: 10.1109/TCSII.2025.3583043.
7. K. Dhanumjaya, M. Sudha, M. N. Giri Prasad, and K. Padmaraju, "Cell Stability Analysis of Conventional 6T Dynamic 8T SRAM Cell in 45nm Technology," International Journal of VLSI Design & Communication Systems (VLSICS), vol. 3, no. 2, pp. 41–51, Apr. 2012.
8. S. R. Sanjana, B. Ramakrishna, Samiksha, R. Banu, and P. Shubham, "Design and Performance Analysis of 6T SRAM Cell in 22nm CMOS and FinFET Technology Nodes," in 2017 International Conference on Recent Advances in Electronics and Communication Technology, IEEE, 2017, doi: 10.1109/ICRAECT.2017.65.
9. B. Satheesh and P. Benakop, "Design of an AAM 6T-SRAM Cell Variation in the Supply Voltage for Low Power Dissipation and High Speed Applications using 20nm FinFET Technology," in Proceedings of the Fifth International Conference on Inventive Computation Technologies (ICICT), IEEE, 2020, pp. 1080–1085.
10. G. Thakral, S. P. Mohanty, D. Ghai, and D. K. Pradhan, "P3 (Power-Performance-Process) Optimization of Nano-CMOS SRAM using Statistical DOE-ILP," in IEEE Computer Society Annual Symposium on VLSI, 2011.

11. P. Meinerzhagen, S. M. Y. Sherazi, A. Burg, and J. N. Rodrigues, "Benchmarking of standard-cell based memories in the sub-VT domain in 65-nm CMOS technology," IEEE Journal on Emerging and Selected Topics in Circuits and Systems, vol. 1, no. 2, pp. 173–182, 2011, doi: 10.1109/JETCAS.2011.2162159.
12. P. Meinerzhagen et al., "A sub-VT 6T SRAM in 28nm FinFET technology," IEEE Journal of Solid-State Circuits, vol. 50, no. 7, pp. 1741–1750, Jul. 2015.
13. M. Ebrahimi, A. Jahanian, and A. Afzali-Kusha, "Read stability enhancement in FinFET SRAM cells," IEEE Transactions on Very Large Scale Integration (VLSI) Systems, vol. 26, no. 5, pp. 915–923, May 2018.
14. T. Song, J. Lee, and S. Kim, "Variation tolerant SRAM bitcell design in FinFET technologies," IEEE Transactions on Electron Devices, vol. 65, no. 3, pp. 1025–1032, Mar. 2018.
15. A. Islam and M. Hasan, "Leakage characterization of 10T SRAM cell," IEEE Trans. Electron Devices, vol. 59, pp. 631–638, 2012.
16. A. Lambora, K. Gupta, and K. Chopra, "Genetic Algorithm - A Literature Review", 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th-16th Feb 2019.
17. M. Alioto, "Energy-quality scalable adaptive VLSI circuits and systems beyond approximate computing," IEEE Journal of Solid-State Circuits, vol. 52, no. 4, pp. 911–924, Apr. 2017.
18. S. Birla, R. K. Singh, and M. Pattnaik, "Static noise margin analysis of various SRAM topologies," IACSIT International Journal of Engineering and Technology, vol. 3, no. 3, pp. 304–309, Jun. 2011, doi: 10.7763/IJET.2011.V3.242.
19. H. Noguchi et al., "Which is the best dual-port SRAM in 45-nm process technology?," Proceedings of IEEE International Conference on IC Design and Technology, pp. 55–58, 2007.
20. T. Kim, J. Liu, and C. Kim, "A voltage scalable 0.26V, 64kb 8T SRAM with Vmin lowering techniques," IEEE Journal of Solid-State Circuits, vol. 47, no. 10, pp. 2295–2306, Oct. 2012.
21. B. H. Calhoun and A. Chandrakasan, "A 256-kb sub-threshold SRAM in 65nm CMOS," IEEE Journal of Solid-State Circuits, vol. 42, no. 3, pp. 680–688, Mar. 2007.
22. K. Takeda et al., "A read-static-noise-margin-free SRAM cell for low-VDD and high-speed applications," IEEE Journal of Solid-State Circuits, vol. 41, no. 1, pp. 113–121, Jan. 2006.
23. J. P. Kulkarni, K. Kim, and K. Roy, "A 160mV robust Schmitt trigger based subthreshold SRAM," IEEE Journal of Solid-State Circuits, vol. 42, no. 10, pp. 2303–2313, Oct. 2007.
24. H. Pilo et al., "A 64Mb SRAM in 32nm high-k metal-gate CMOS technology with 0.7V operation," IEEE Journal of Solid-State Circuits, vol. 44, no. 1, pp. 167–178, Jan. 2009.
25. A. Bhavnagarwala, X. Tang, and J. Meindl, "The impact of intrinsic device fluctuations on CMOS SRAM cell stability," IEEE Journal of Solid-State Circuits, vol. 36, no. 4, pp. 658–665, Apr. 2001.
26. R. S. Chalapathi and C. Kumar, "Optimized Design and Performance Analysis of Multi-Transistor SRAM Cells for Low-Power VLSI Applications in 45nm CMOS Technology", Journal of Systems Engineering and Electronics (ISSN: 1671-1793), vol. 35, no. 9, 2025.
27. Q. Zheng, Y. Wu, C. Zhao, and J. Zhou, "Optimization of 6T-SRAM Cell Based on CNN-Informed NSGA-II with Consideration of Parasitic Resistance", MDPI Electronics, vol. 14, no. 20, art. 4002, 2025, doi: 10.3390/electronics14204002.
28. J. Kennedy and R. Eberhart, "Particle Swarm Optimization," in Proc. IEEE Int. Conf. Neural Networks, 1995, pp. 1942-1948.
29. R. Storn and K. Price, "Differential Evolution - A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces," Journal of Global Optimization, vol. 11, pp. 341-359, 1997.
30. M. Dorigo, V. Maniezzo, and A. Colorni, "Ant System: Optimization by a Colony of Cooperating Agents," IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 26, no. 1, pp. 29-41, 1996.
31. D. E. Goldberg, Genetic Algorithms in Search, Optimization, and Machine Learning. Boston, MA: Addison-Wesley, 1989.
32. N. Weste and D. Harris, CMOS VLSI Design: A Circuits and Systems Perspective, 4th ed. Boston, MA: Pearson, 2010.

DOI - 10.68648/ijte-q-2026-49-03-017

Bending analysis of Compact geometry Monopole Antenna for Wearable Applications

Ambresh P. A.
Department of Applied Electronics
Gulbarga University
Kalaburagi, Karnataka
✉ ambreshpa@rediffmail.com

ABSTRACT

This work describes a compact monopole antenna design including bending analysis. The antenna has a variable return loss of -41dB and a bandwidth of 1422MHz for the lower microwave band resonance of 4.52GHz. The bending analysis along the X, Y axis planes are performed at different phases. V2X high-speed wireless communication, 5G WLAN, 5G-NR frequencies, fifth-generation Sub-6GHz (5G) smartphone apps, and C-band wearable applications are all intended uses for the designed antenna. With a thickness of 24 x 29 x 0.254 mm³, the substrate exhibits compact geometry. Through simulations employing tissue that resembles a human arm model, a gain value of 3.64dBi with SAR values that remain within the allowable range are also attained.

KEYWORDS *: DWearable, 5G, Low weight, Antenna, Compact, Flexible.*

INTRODUCTION

A growth is found in the requirement of wearable devices that can track bodily functions due to the shrinking of electronic systems [1], making wearable health management systems a very desirable area for researchers to study. This is because, when integrated into smart clothing, wearable technology can perform well enough to replace a number of medical devices [2].

LITERATURE REVIEW

Researchers created a range of compact antennas with sufficient performance in order to implement such technology [3]. However, a number of practical antenna design requirements, including compact size, low weight, power consumption, and flexible structure can be used in wearable devices for biomedical applications [4]. Thus, the use of many microstrip antennas, which are among the most sought types for wearable applications when installed in near-closeness to the human body [5].

In addition to wearable devices, flexible antenna designs are also proposed for millimeter wave applications, wireless power transmission, and 5G wireless communications to work with next-generation flexible electronics. A flexible broadband antenna based on Rogers RO4835T is proposed in [6] communication systems in 5G sub-6GHz band. Hexagonal split-ring resonators are recommended because the radiating element offers broadband capability for sub-6 GHz, preventing interference problems in the frequency range. Without sacrificing antenna performance, these resonators may be circumscribed independently.

For systems equipped with RFID tags, a design for a flexible, transparent antenna that operates across multiple bands is given in [7]. This design demonstrates wireless power transfer in both radiation field scenarios, operating at resonant frequencies of 13.56MHz and 2.45GHz, respectively. A flexible antenna which consists of a rectangular patch embedded within a flexible PDMS substrate is studied in [8]. As challenging environments such as channel imbalance and body attenuation effects draw growing attention, wearable antennas must operate effectively within particular frequency bands to be used in different communication based wireless protocols. The prevalence of these frequency bands varies, including 2.4GHz and 5.8GHz-ISM bands, UWB (3.1–10.6 GHz), and up-and-coming 5G bands (24–40 GHz) [9-13]. Architectures are chosen according to the use case, yet wideband and multiband

designs are commonly adopted so that several functions—such as GPS, Bluetooth, LTE cellular connection, and Wi-Fi which can operate at the same time. This allows the antenna to adapt with ease across various communication protocols.

Considerable focus was given to employing flexible and semi-flexible materials during fabrication of wearable nodes[14]. Thanks to major breakthroughs in wearable antenna technologies, their application in implantable devices has grown remarkably [15]. Nevertheless, antennas with low frequencies while keeping them miniaturized for wearable devices, all without compromising radiation efficiency and gain, remains among the most pertinent challenges in present days [16]. To account this problem, a variety of methods are pursued to shrink wearable antennas, including reactive loads [17], high-permittivity materials [18], vias [19], shorting posts [20], and fractal geometries that lengthen the current path [21]. The antenna functioning with close attention to how different its positions particularly the remoteness between the fabric material and skin are studied in [25]. For wearable applications, particularly with flexibility nature of substrate along with its design, structure are thoroughly examined in [26], focusing on their role in WBAN and medical-health technology, with highlighting the need to sustain optimal performance under human-body contact and also in mechanical deformation condition. The paper [27] discusses the novel materials employed in wearable electronics devices.

The objective reported in this paper is to create a simpler antenna that is compact in size, offers a reasonably wide bandwidth, and maintains acceptable Specific Absorption Rate (SAR) levels for wearable uses. The wearable antenna's design is detailed in section II. In section III, the findings are discussed, with conclusion in section IV, followed by acknowledgments and references at the end.

WEARABLE ANTENNA DESIGN

A compact microstrip flexible wearable antenna featuring a circular patch engraved I-shape structure in the centre of the radiating patch, alongside an outer circular cut patch next to it, excited using a line-feed method with a partial copper acting as ground on the other side of substrate is shown in fig.1. The suggested antenna employs Rogers RT Duriod 5880 as the substrate, and the radiating patch is constructed using copper with a thickness of 0.035 microns.

The designing of circular patch antenna involves determining the radius and effective radius according to the operating frequency, dielectric constant, and substrate height. The equations comprise [22].

$$a = \frac{F}{\left\{1+\frac{2h}{\pi \epsilon_r F}\left[\ln\left(\frac{\pi F}{2h}\right)+1.77726\right]\right\}^{\frac{1}{2}}} \quad (1)$$

where,

$$F = \frac{8.791\times10^{9}}{f_r\sqrt{\epsilon_r}} \quad (2)$$

and determining effective dielectric constant for fringing estimation.

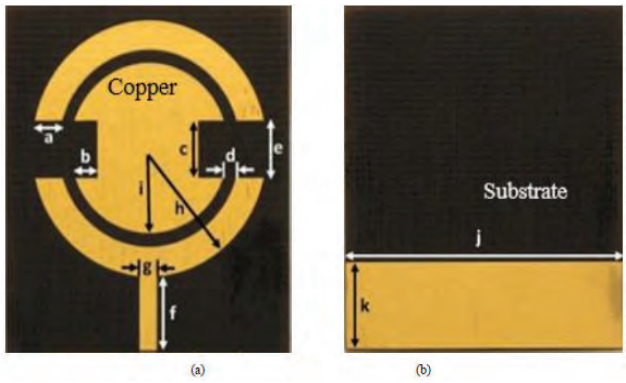


Fig.1. Design of antenna. a) Front view, b) Back view

The antenna layout illustrated in fig. 1 features an external circular patch with a radius of h=11.6 mm on a Rogers RT 5880 flexible substrate (with εr = 2.2, tanδ = 0.0009, Hs = 0.254 mm), which is energized by a 50-Ω coaxial probe. This antenna includes an I-shaped inner circular patch at the center, which has a radius of i=9.2mm for the radiation structure. The yellow shade indicates the copper layer, where as brown area represents the substrate layer in figure 1 illustrating the front and back layers. The different parameters employed in the antenna design along with their sizes are a= 2.62mm, b= 2.61mm, c= 5.1mm, d= 1.02mm, e= 5.1mm, f= 5.60mm, g= 1.52mm, h= 11.4m, i= 9.1mm, j= 26mm, k= 4.73mm.

BENDING ANALYSIS RESULTS

Simulation study is carried out on the proposed antenna design. For this purpose, the S_{11} (return loss) spectra and radiation patterns of the proposed antenna were measured across resonant frequency band under different bending conditions in simulation environment. Afterward, the same set of design was repeated with the antenna positioned near the phantom arm tissue model. Measurements of the proposed antenna were taken over the frequency range of 3 GHz to 6 GHz. For performing simulations, the CST Studio Suite software is used, and the antenna results, encompassing the reflection coefficient of bending examination of the antenna, Specific Absorption Rate (SAR), and VSWR are incorporated in this section. The S_{11} response shown in Fig.2 illustrates stable impedance matching and a wide impedance bandwidth of 1422 MHz, achieving a return loss of -41dB at 4.52GHz. The simulated bandwidth ranges from 4.053 GHz to 5.475 GHz. The suggested antenna for wearable use is created with an emphasis on compact devices such as smartwatch applications or for the case of wearable clothes or devices. When this kind of device is used on arm, it experiences a degree of conformability or bending nature. So, the bending characteristic is studied and the antenna's performance in this case should stay near-to steady, or results should be similar to those of the antenna in a typical situation without any bending situations.

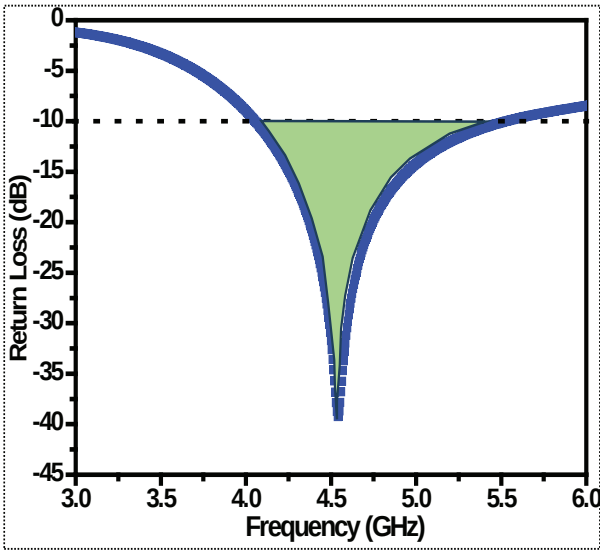


Fig. 2. Reflection Coefficient of designed patch [23]

Consequently, the suggested antenna is curved along the x- and y-axis. Simulation results show that, the antenna is curved in the x-direction at a 90◦ angle and has a bending radius in steps of 20 mm. Since the antenna is designed for hand-worn use, a tighter bending radius was considered. Fig. 3 (a) illustrates the curved design of the wearable antenna from an isometric perspective, while the top and bottom perspectives are also displayed respectively.

Correspondingly, fig. 4 illustrated the proposed antenna under conformal conditions along the y-axis at a 90◦ angle, with a bending radius of 20 mm per step beginning at 40 mm, applying the same bending angle and radius as those utilized for the x-axis. Along both axes, the bent antenna's reflection characteristics match closely those of the simulated antenna that has no bending. As the simulation results in Fig. 4 show, the proposed antenna, when bent along both axes, the antenna functions across the 3.8 GHz to 5.3 GHz range (wideband) suitable for wearable applications and Sub-6 GHz 5G applications.

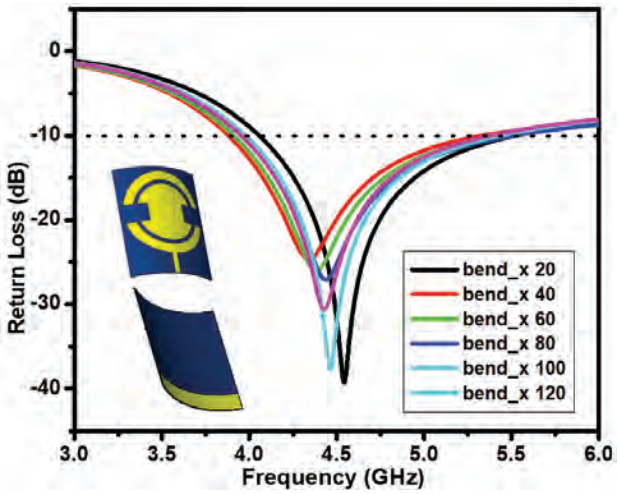


Fig. 3. Reflection Coefficient of antenna bend on x-axis

When the antenna was subjected to the bending effects, it exhibited insignificant changes in terms of the S_{11} aspect, as seen in Figures 3 and 4. This variation in the frequency and bandwidth results from the current motion on the patch structure and the side cuts along the x-axis, and also from the line feed point.

Because the proposed study involved the use of wearable devices, a SAR investigation was also carried out to examine radiation leaking from the antenna into human tissue. This examines the consequences

of radiation absorbed within the human arm. For the simulation, the input power level was set at 1 mW. The proposed antenna design in order to achieve minimal leakage into human tissue is achieved by placing a feed port near to the feeding line. Here, the antenna was positioned close to the human arm mimic phantom, at a distance of 0 mm. It should be noted that, a bending effect of 90◦ as flat case was taken into account. The Specific Absorption Rate (SAR) measures how much radiofrequency energy is taken in by human tissue. It is determined by averaging over a specific mass of 1 gram. The SAR limit in US is 1.6 W/kg for 1 gram of tissue, which corresponds to 2 W/kg in Europe for 10 grams of tissue.

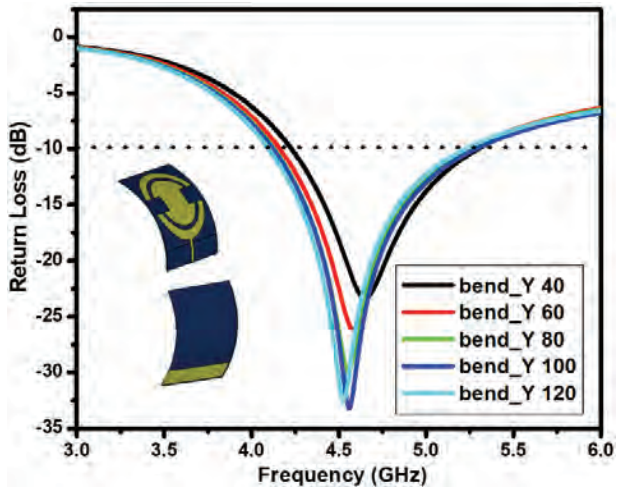


Fig. 4. Reflection Coefficient of antenna bend on Y-axis



Fig. 5. (a)Human tissue layers, (b) layered colours and (c) SAR at 1g. [23]

The SAR value is highly important, as it demonstrates how back radiation affects human tissue. If it surpasses the recommended limit set by the FCC and ICINPR, it will damage the tissues of the human body. To examine this, a region of skin, fat, and muscle keeps dimensions constant at 50 × 50 mm². The thickness of these layers, namely skin, fat and muscle, is kept at 2 mm, 3 mm, and 8 mm respectively. The antenna is modeled and evaluated on the skin, while the dielectric constant stays at 41.4 with thermal conductivity measuring 0.88 S/m, representing human biological tissues, are employed in the simulations as given in Fig.5(a). The fig. 5 (b) shows the different layers of the tissue considered for the study. The specific absorption rate (SAR) analysis using this confirms safe operation at 4.52 GHz, with a 1-g measurement of 1.19 W/kg for the arm-tissue model, well below the IEEE/IEC limits of 1.6 W/Kg, as shown in Fig. 5(c). The antenna also shows enhanced gain, with a slight improvement in the on-body case (i.e., 3.64 dBi simulated), likely due to favorable interaction with the body surface modeled as phantom layers in the design. Figure 6 illustrates the VSWR plotted against frequency, revealing that the antenna attains a VSWR value below 1.5, which indicates improved impedance matching and reduced power being reflected back toward the source.

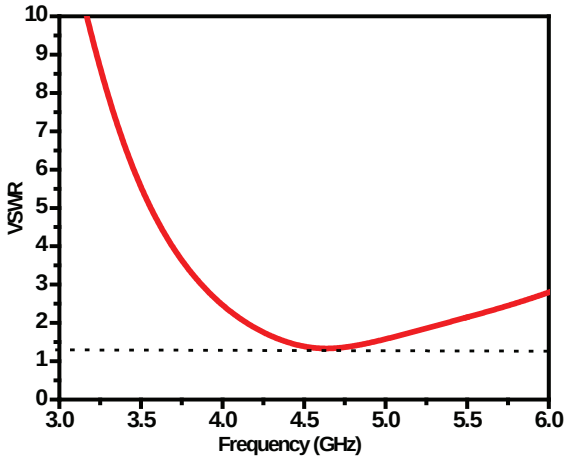


Fig. 6. Plot of VSWR Vs frequency for the proposed design

Figure 7a, b shows the 2D and 3D radiation characteristics of the designed antenna depicting that the antenna exhibited the nearly broadsided radiation pattern at the desired resonance frequency.

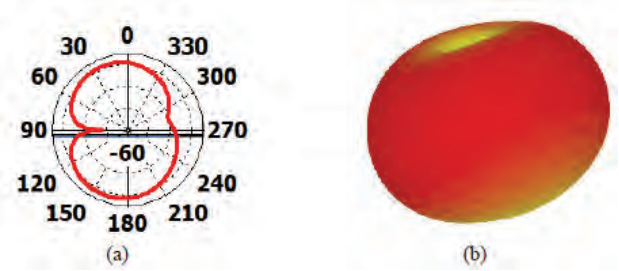


Fig. 7 a)2D, b)3D radiation characteristics

USES OF WEARABLE ANTENNAS

The demand from users for wearable devices has expanded swiftly with the emergence of modern monitoring systems. In particular, their presence within the healthcare system is growing quickly, which is expected to benefit society positively. Physiological conditions serve as vital health indicators, and tracking them could effectively permit the early identification of any type of disease. Smart wearable applications would lessen significant health difficulties, including disease management and expense. Typically, clinical equipment used for measurement is not wearable; therefore, researchers have concentrated on distinctive smart configurations that would make wearable devices possible with a low profile, robustness, and strong performance. The IEEE 802.15 group has acknowledged standard applications for body-centric communications [24].

CONCLUSION

This study presents a compact wideband wearable antenna that achieves a gain of 3.64dBi, shows a bidirectional radiation pattern, and offers a broad impedance bandwidth of 1422MHz at 4.52GHz, which makes it suitable for wearable applications. Using CST simulation studio, the SAR performance of the proposed antenna was assessed and determined to be 1.19W/Kg, a value considerably lower than the standard 1.6W/Kg. Future work could be carried out with a metasurface or AMC in order to achieve improved bandwidth and gain results.

ACKNOWLEDGEMNTS

We express our gratitude to Vision Group on Science & Technology (VGST), Karnataka Science & Technology Promotion Society (KSTePs), Govt of Karnataka, (Grant no: VGST/ KSTePs: GRD.1218) for channeling the funds through the KFIST scheme, sanctioned in July 2025, to support this work.

REFERENCES

1. M.N.N Alaukally, T.A.Elwi, D.C Atilla (2021), “Miniaturized flexible metamaterial antenna of circularly polarized high gain-bandwidth product for radio frequency energy harvesting”,. Int. J. Commun. Syst. e5024.
2. R.K Abdulsattar, T.A.Elwi, Z.A.A Hassain (2021), “A New Microwave Sensor Based on the Moore Fractal Structure to Detect Water Content in Crude Oil”, Sensors, Vol. 21, pP.7143.
3. D. Ali, T. Elwi, S. Özbay (2021), “Metamaterial-based Printed Circuit Antenna for Biomedical Applications”, Avrupa Bilim Ve Teknol. Derg., Vol. 26, pP12–15.
4. J Lmako (2000), “Specific Absorption Rate (SAR) Testing, from a Test Laboratory Perspective”, No. 211 7901; Radio Frquency Investigation Ltd.: Basingstoke, UK.
5. N. Kumaran, K.Arunachalam, “A Wideband non-Resonant FSS with Finite Number of Unit Cells for Mobile Phone SAR Reduction”, In Proceedings of the International Conference on Electromagnetics in Advanced Applications (ICEAA), Cairns, QLD, Australia during 19–23 September 2016, pp. 901–904.
6. A. Zaidi, W.A. Awan, N. Hussain, A. Baghdad (2020), “A wide and tri-band flexible antennas with independently controllable notch bands for sub-6-GHz communication system”, Radio Engineering. Vol. 29, pP44–51.
7. A.S.M. Sayem, D. Le, R.B.V.B. Simorangkir, T. Bjorninen, K.P. Esselle, R.M. Hashmi, M. Zhadobov (2020), “Optically transparent flexible robust circularly polarized antenna for UHF RFID tags”, IEEE Antenn. Wireless Propag. Lett. Vol. 19, pP2334–2338. https://doi.org/10.1109/LAWP.2020.3032687.
8. S.S. Dhanabalan, S. R, A K. Madurakavi, R M Thirumurugan, S.R. Avaninathan, M.F. Carrasco (2022), “Flexible compact system for wearable health monitoring applications”, Comput. Electr. Eng. Vol. 102.
9. Monti G., Corchia L., De Benedetto E., & L Tarricone (2016), “Wearable logo-antenna for GPS–GSM- based tracking systems”, IET Microwaves, Antennas & Propagation Research Article, Vol.10, No.12, pP.1332–1338.2016. https://doi.org/10.1049/iet-ap.2015.0774.
10. Zong-zuo Y. & G.-Z Guo (2017), “Improvement of positioning technology based on RSSI in ZigBee networks” Wireless Personal Communications, Vol. 95, No. 3, pP. 1–20. https://doi.org/10.1007/s11277-016-3860-1.
11. Pei R., Leach M. P., Lim E. G, Wang,Z., Song C., Wang J., Zhang W., Jiang Z. & Y. Huang (2020), “Wearable EBG-backed belt antenna for smart on-body applications”, IEEE Transactions on Industrial

Informatics,Vol.16, No.11, pP. 7177–7189. https://doi.org/10.1109/TII.2020.2983064.

12. Wang L., Jianguo Y., Xie T. & B. Kun (2021), "A novel multiband fractal antenna for wireless application", International Journal of Antennas and Propagation, (Article ID 9926753), 9.https://doi.org/10.1155/2021/9926753.

13. Varma S., Sharma S., , John M., , Bharadwaj R, Dhawan A & S. K. Koul (2021), "Design and performance analysis of compact wearable textile antennas for IoT and body-centric communication applications", International Journal of Antennas and Propagation, (Article ID 7698765). https://doi.org/10.1155/2021/7698765.

14. Elwi, T.A.; Imran, A.I.; Alnaiemy (2015), Y. A Miniaturized Lotus Shaped Microstrip Antenna Loaded with EBG Structures for High Gain-Bandwidth Product Applications. Prog. Electromagn. Res. C, Vol.60, pP.157–167.

15. Elwi, T.A. A Miniaturized Folded Antenna Array for MIMO Applications. Wirel. Pers. Commun. 2008, 98, 1871–1883.

16. Imran, A.I.; Elwi, T.A.(2017), A Cylindrical Wideband Slotted Patch Antenna Loaded with Frequency Selective Surface for MRI Applications. Eng. Sci. Technol. Int. J. 2017, Vol. 20, pP.990–996.

17. Abdulmjeed, A.; Elwi, T.A.; Kurnaz, A.S.(2021), Metamaterial Vivaldi Printed Circuit Antenna Based Solar Panel for Self-Powered Wireless Systems. Prog. Electromagn. Res. M, Vol. 102, pP.181–192.

18. Imran, A.I.; Elwi, T.A.; Salim, A.A.J (2021), On the Distortionless of UWB Wearable Hilbert-Shaped Metamaterial Antenna for Low Energy Applications. Prog. Electromagn. Res. M, Vol.101, pP.219–239.

19. Al-Dulaimi, Z.; Elwi, T.A.; Atilla, D.C.(2020), Design of a Meander Line Monopole Antenna Array Based Hilbert-Shaped Reject Band Structure for MIMO Applications. IETE J. Res. pP. 1–10.

20. Li, J. Computer Aided Modeling and Simulation of Cooling Fan System. In Proceedings of the 2011 International Symposium on Information Engineering and Electronic Commerce (IEEC2011), San José, CA, USA, 5–9 June 2011.

21. Elwi, T.A.; Al-Rizzo, H.M.; Rucker, D.G.; Khaleel, H.R (2010), Effects of twisting and bending on the performance of a miniaturized truncated sinusoidal printed circuit antenna for wearable biomedical telemetry devices. AEU-Int. J. Electron. Commun. Vol.13,pP.1–12.

22. Balains, C. A. "Antenna Theory Analysis and Design". 3rd Edition John Wiley & Sons. New Jersey, 2005.

23. Ambresh P A, "Wideband Flexible Monopole Antenna with Compact Geometry and Partial Ground for Wearable Applications", Second International Conference on Next Generation Communication and Information Processing (INCIP2026), CUK Kalaburagi during Aug. 20-21,2026.

24. Rais, N.H.M.; Soh, P.J.; Malek, F.; Ahmad, S.; Hashim, N.; Hall, P, A review of wearable antenna. In Proceedings of the 2009 Loughborough Antennas & Propagation Conference, Loughborough, UK, 16–17, 2009; pP. 225–228.

25. Lundquist, J.; Linkous, L.; Piper, M.E.; Sickey, Z.; Zimmet, K.; Mendoza, I.; Suresh, S.; Topsakal, E. (2024), Textile-based inkjet-printed RFIDs: Exploring wearable antennas in the real world [Bioelectromagnetics]. IEEE Antennas Propag. Magv, Vol.66, pP50–62.

26. Singh, S.; Mishra, R.; Kapoor, A.; Singh, S. (2025), A Comprehensive Review and Analysis of the Design Aspects, Structure, and Applications of Flexible Wearable Antennas. Telecom, Vol. 6, pP3.

27. Yang, M.; Ye, Z.; Ren, Y.; Farhat, M.; Chen, P.-Y. (2024), Materials, Designs, and Implementations of Wearable Antennas and Circuits for Biomedical Applications: A Review. Micromachines, Vol.15, PP26.

DOI - 10.68648/ijte-q-2026-49-03-018

Massification of Technical Education in India: Implications for Quality and Academia–Industry Alignment

Keshav Kant
Retired Professor
Department of Mechanical Engineering
Indian Institute of Technology Kanpur
✉ keshav@iitk.ac.in

ABSTRACT

Over the past three decades, India has experienced one of the world's largest expansions of higher and technical education, substantially increasing institutional capacity and access to engineering education. While this massification has broadened educational opportunities and supported socio-economic development, improvements in educational quality have not always kept pace with quantitative growth. This review examines the implications of massification for educational quality, faculty development, research ecosystems, laboratory infrastructure, graduate employability, and academia–industry alignment. It argues that rapid institutional expansion, particularly during the period of intensive private-sector growth, often outpaced the development of the academic ecosystems required to sustain high-quality technical education. The review also examines how Artificial Intelligence is reshaping engineering curricula, pedagogical practices, and future graduate competencies. Drawing upon national policy documents, international literature, and recent developments in engineering education, the paper argues that the next phase of reform should emphasize institutional differentiation, continuous faculty development, curriculum modernization, stronger academia–industry collaboration, sustainable funding mechanisms, and outcome-based quality assurance. Having successfully expanded access to technical education, India must now consolidate these gains by strengthening educational quality, innovation capacity, graduate employability, and workforce relevance.

KEYWORDS *: Technical education; Massification; Engineering education; Academia–industry alignment; Artificial Intelligence.*

INTRODUCTION

The expansion of higher education from elite to mass participation has been recognized as one of the defining transformations of modern societies. Trow's seminal framework describes this transition as a progression from elite to mass and ultimately universal access, accompanied by profound changes in institutional missions, governance, curriculum, and quality assurance [1]. India's higher and technical education system provides one of the world's largest contemporary examples of this transformation.

India's higher and technical education system has undergone an unprecedented phase of expansion over the past two decades [2,4]. Driven by demographic pressures, rising aspirations of a growing middle class, and policy emphasis on increasing the Gross Enrolment Ratio (GER), the country witnessed a rapid proliferation of universities, engineering colleges, and technical institutions. This expansion was considered essential for harnessing India's demographic dividend and supplying the skilled human resources required for a rapidly growing economy [2,3]. The rapid expansion of technical institutions was not always accompanied by proportional investments in laboratory infrastructure, faculty development, technical support staff, and institutional governance. Consequently, capacity expansion often outpaced the development of academic ecosystems necessary to sustain educational quality. This imbalance between quantitative expansion and qualitative development has emerged as one of the defining challenges of India's technical education system.

Between the mid-2000s and late 2010s, the number of engineering institutions increased sharply, with private participation accounting for a substantial share of this growth [4,5]. Enrolment capacity expanded correspondingly, extending access to technical education across a wider geographical and socio-economic spectrum. In quantitative terms, the system achieved an unprecedented scale. The long-term strength of India's technological capability and its aspirations for innovation-led development will depend on how effectively this expanding technical education system reconciles growth with quality.

Recent reforms, particularly the National Education Policy (NEP) 2020, seek not only wider access, but also improvements in quality, equity, multidisciplinary education, research and innovation, and graduate employability [2].

However, numerical expansion alone cannot guarantee educational quality [4–6]. The institutional capacity required to sustain standards—qualified faculty, research culture, laboratory infrastructure, effective governance, and meaningful industry engagement—often evolves more slowly than physical infrastructure or sanctioned intake. As a result, concerns have emerged regarding uneven academic standards, graduate employability, and the depth of industry–academia collaboration [4].

At its core, however, the fundamental responsibility of technical education remains the imparting of sound knowledge, the development of practical competence, and the preparation of graduates capable of contributing productively to the economy and society. The present challenge is no longer one of expansion alone, but of consolidation, alignment, and qualitative strengthening. The transition from massification to meaningful educational quality has thus become one of the defining policy challenges confronting Indian technical education. Against this background, this review examines the evolution of technical education in India, evaluates the implications of massification for educational quality, faculty development, research capacity, graduate employability, and academia–industry alignment, and discusses how emerging technologies, particularly Artificial Intelligence, are reshaping engineering education and the competencies expected of future graduates [9–17]. The paper concludes by identifying policy and institutional measures needed to reconcile expansion with excellence.

THE EXPANSION PHASE: GROWTH, ACCESS AND STRUCTURAL IMBALANCES

Policy Context and Growth in Enrolment

The rapid expansion of technical education in India must be viewed in the context of broader policy objectives aimed at increasing access to higher education [2, 3]. Over the past two decades, the Gross Enrolment Ratio (GER) in higher education has risen from low double-digit levels in the early 2000s to nearly 28–29% in the early 2020s, reflecting both demographic pressures and deliberate policy initiatives to widen participation [2, 4].

From the perspective of Trow's framework, India has undergone a transition from an elite system serving a relatively small proportion of the population to a mass system of higher and technical education. This transformation has greatly improved educational access, but has simultaneously created new challenges related to quality assurance, institutional differentiation, graduate employability, and resource allocation [1].

Engineering and technical education formed a central component of this expansion, given their perceived role in supporting industrial growth, infrastructure development, and technological advancement [2, 3]. The number of universities more than doubled during this period, and the expansion of engineering institutions was particularly pronounced between 2005 and 2015 [4,5]. This period marked the most intensive phase of massification in Indian technical education, with expansion occurring at a pace unprecedented in the country's higher education history.

A substantial proportion of these new institutions emerged in the private sector, which today accounts for the majority of engineering colleges in the country. This private-led expansion enabled rapid scaling without placing a disproportionate fiscal burden on the public exchequer and increased access in several semi-urban and rural regions. However, the expansion of physical capacity was not always matched by comparable investments in laboratory infrastructure, trained technical staff, and opportunities for meaningful hands-

on learning. Consequently, practical competencies and graduate readiness varied considerably across institutions. Similarly, the rapid expansion generated demand for qualified faculty that often exceeded supply, resulting in variability in teaching quality, mentoring, and laboratory instruction.

Intake Expansion and Market Saturation

Approved intake capacity in engineering programmes grew sharply during the expansion phase, particularly between 2005 and 2015, at times outpacing actual enrolment demand [5]. In several states—particularly those that experienced early private-sector growth—seat occupancy rates began to decline after an initial surge of enthusiasm.

This created institutional stress, with some colleges operating below capacity and others eventually closing or consolidating. The episode illustrates an important structural reality: while policy incentives can stimulate rapid institutional creation, student demand, institutional quality, and labour market absorption do not always expand at the same pace [4,6].

A structural paradox emerged. National capacity expanded significantly, yet institutional viability and academic standards in many colleges came under strain due to under-enrolment [4,5]. Persistent under-enrolment in several engineering disciplines has resulted in institutional consolidation and the closure of a number of engineering colleges in different parts of the country, indicating that expansion has entered a phase of rationalization.

Regional and Tiered Imbalances

The geographic distribution of engineering institutions has been uneven [4]. Southern and western states witnessed dense concentrations of private engineering colleges during the expansion phase, while parts of eastern and central India expanded more gradually.

Expansion also occurred unevenly across institutional tiers. Premier public institutions—such as the Indian Institutes of Technology and National Institutes of Technology—expanded cautiously while continuing to invest in faculty development and research infrastructure [5,6]. In contrast, many newly established institutions operated with limited research ecosystems and relied predominantly on teaching-focused models.

This asymmetry contributed to the emergence of a stratified technical education system characterized by significant differences in faculty quality, research capability, infrastructure, graduate outcomes, and industry engagement [4,6]. These structural imbalances created many of the quality-related challenges that are examined in the following section.

QUALITY UNDER STRAIN: FACULTY, RESEARCH AND INFRASTRUCTURE

Teaching Mandate and Academic Responsibility

Sustaining academic quality in technical education depends fundamentally on faculty competence, research engagement, laboratory infrastructure, and academic governance [4–6]. The primary mission of technical institutions remains the delivery of high-quality engineering education and the preparation of competent graduates capable of contributing effectively to the economy and society. Research excellence and external funding are important dimensions of academic progress—particularly in premier institutions—but they should complement rather than displace the central teaching mission. A balanced integration of teaching, research, innovation, and professional engagement is therefore essential for maintaining academic standards and ensuring the long-term relevance of technical education [6,7].

Faculty Availability and Academic Stability

One of the most persistent challenges confronting technical education in India has been the availability of qualified faculty [4–6]. The rapid establishment of new engineering institutions generated demand for faculty that exceeded the available supply. Consequently, many institutions relied on early-career graduates with limited teaching and industrial experience, affecting curriculum delivery, laboratory instruction, student mentoring, and industry interaction. The primary mission of technical education is to develop conceptual understanding, analytical ability, practical competence, and professional confidence rather than merely preparing students to succeed in university examinations.

While such arrangements ensured operational continuity, they often limited the development of stable academic cultures, long-term mentoring relationships, and sustained research programmes. Variability in

compensation structures, career progression, and faculty development opportunities further contributed to uneven academic capacity, particularly across segments of the private sector [4–6].

Research Concentration and Institutional Heterogeneity

Research intensity across institutions has remained highly uneven [6,7]. A relatively small cluster of nationally funded institutions accounts for a disproportionate share of publications, patents, and externally funded projects. Many teaching-oriented colleges have limited research output, owing to constrained funding, heavy teaching loads, and modest industry collaboration [6,7].

Consequently, the integration of research with undergraduate teaching—a hallmark of globally competitive technical universities—remains limited across large segments of the system, thereby constraining innovation, experiential learning, and the development of research-oriented engineering graduates [6,7].

Infrastructure and Quality Assurance

In recent years, accreditation has increasingly shifted towards Outcome-Based Education (OBE), emphasizing measurable graduate attributes, continuous improvement, curriculum relevance, and stakeholder participation. Accreditation and quality assurance mechanisms—particularly those administered by the National Assessment and Accreditation Council (NAAC) and the National Board of Accreditation (NBA)—have sought to address these concerns through periodic evaluations and outcome-based assessment frameworks [18]. While accreditation has significantly strengthened institutional accountability and promoted continuous quality improvement, the effectiveness of its implementation has varied across institutions. In some cases, greater emphasis has been placed on documentation and procedural compliance than on deeper academic transformation in teaching, research, and industry engagement [4,18].

Laboratory infrastructure constitutes a defining component of engineering education. Although physical infrastructure expanded rapidly during the massification phase, the availability of modern equipment, maintenance support, trained technical staff, and opportunities for experiential learning has varied considerably across institutions [4]. Such disparities limit opportunities for laboratory work, workshops, design projects, and industrial training, thereby affecting practical competence, engineering judgement, innovation, problem-solving ability, and graduate readiness for industry [4,18].

ACADEMIA–INDUSTRY ALIGNMENT: SKILL MISMATCH AND INNOVATION GAP

The Employability Debate

One of the most widely discussed consequences of the rapid expansion of technical education has been the gap between graduate competencies and industry expectations [3,17,20]. Employers across sectors have expressed concerns regarding practical skills, problem-solving ability, interdisciplinary exposure, workplace readiness, and the ability to adapt to rapidly evolving digital technologies and AI-assisted engineering environments [10–17]. The combined effects of uneven academic quality, inadequate practical training, and limited industry exposure have contributed to persistent concerns regarding graduate employability [3,17,20]. These concerns are reflected in national employability assessments, such as the India Skills Report, as well as in global workforce analyses including the World Economic Forum's Future of Jobs Report [17,20].

Curriculum Responsiveness and Technological Acceleration

Although mechanisms for periodic curriculum revision exist, technological transformation in sectors such as artificial intelligence, semiconductor design, electric mobility, advanced manufacturing, renewable energy systems, and data-driven engineering has accelerated rapidly in recent years [8,17].

The emergence of generative artificial intelligence, digital twins, Industry 4.0, robotics, and intelligent manufacturing systems has further shortened the life cycle of technical knowledge, requiring curricula to become more agile and continuously updated rather than relying solely on periodic curriculum revision cycles [9–16].

Many nationally recognized institutions follow structured curriculum review cycles; however, the pace

of technological change increasingly challenges the adequacy of periodic curriculum revisions alone. Such institutional differentiation and responsiveness align with broader global trends in higher education reform [3,8,16].

Consequently, curriculum design must evolve from periodic revision towards a more continuous, flexible, and industry-responsive process capable of accommodating emerging technologies and rapidly changing workforce requirements [2, 9–17].

Internships, Apprenticeships and Practical Exposure

Where structured internships, apprenticeships, and industry-linked training opportunities are limited, students may graduate with sound theoretical grounding, but insufficient practical experience [2,3,8]. This gap becomes more pronounced in sectors that require familiarity with shop-floor processes, industrial safety practices, and operational troubleshooting. Industrial internships, laboratory-based learning, capstone design projects, and multidisciplinary team projects play a critical role in developing engineering judgement, professional confidence, and the ability to translate theoretical knowledge into practical solutions [2,18,19]. Comparable challenges are also evident in many diploma and vocational institutions, where constraints in laboratory infrastructure and practical training continue to affect skill development and workplace readiness [4,18]. The uneven distribution of technology transfer offices, incubation centres, innovation hubs, and start-up ecosystems further reflects disparities in institutional–industry integration [7,19].

Pedagogical Design and Soft Skills

Industry expectations extend beyond technical competence. Communication skills, interdisciplinary teamwork, adaptability, and collaborative problem-solving increasingly influence recruitment decisions [17].

The alignment challenge is therefore not merely one of curricular content, but also of pedagogy, assessment, and institutional culture. Without sustained collaboration through curriculum co-design, joint research, faculty immersion programmes, industrial internships, and long-term industry partnerships, structural gaps between technical education and industry requirements are likely to persist [3,16,19].

Innovation Ecosystem and Technology Transfer

Innovation in engineering education extends beyond classroom instruction to include research commercialization, start-up incubation, technology transfer, intellectual property creation, and industry-sponsored research [7,19]. While several leading institutions have developed vibrant innovation ecosystems supported by incubation centres and industry partnerships, many teaching-oriented institutions continue to provide limited opportunities for prototype development, technology commercialization, and entrepreneurial engagement. Strengthening these institutional ecosystems is essential for improving academia–industry alignment and supporting India's transition towards an innovation-driven economy [2,7,19].

Artificial Intelligence: Emerging Implications for Quality and Academia–Industry Alignment

Artificial Intelligence (AI), particularly Generative AI, is emerging as a transformative force in engineering education [9-16]. Unlike earlier technological advances such as computer-aided design (CAD), simulation software, and internet-based learning—which primarily enhanced existing engineering practices—Generative AI can generate text, computer code, engineering designs, and analytical solutions, fundamentally altering the teaching–learning process and the competencies expected of engineering graduates [9–12]. Consequently, AI should be viewed not merely as another digital tool, but as a catalyst for curriculum redesign, pedagogical innovation, and institutional transformation [11,12,15].

The rapid adoption of AI across engineering industries is compelling higher education institutions to re-examine curricula, assessment methods, and learning outcomes [11–17]. Future engineering graduates will require not only sound disciplinary knowledge but also AI literacy, data interpretation skills, critical thinking, ethical awareness, and the ability to evaluate and validate AI-generated outputs [9,11,12,15]. Engineering fundamentals, design capability, laboratory experience, and professional judgement will remain indispensable, while AI will increasingly serve as an enabling technology rather than a substitute for engineering competence.

The successful integration of AI into technical education depends critically on faculty preparedness [9,14]. Although there is considerable interest and institutional intent to incorporate AI into engineering curricula, faculty preparedness and training have yet to reach the level required for its effective implementation [14]. Sustained faculty development, curriculum innovation, and closer collaboration with industry will therefore be essential if AI is to strengthen educational quality and improve academia–industry alignment rather than merely introduce another technological layer into engineering education [10,14]. Artificial Intelligence should therefore be viewed as a catalyst for transforming engineering education, provided that institutions adopt appropriate pedagogical practices, ethical safeguards, and regulatory frameworks [11,12].

STRUCTURAL AND REGULATORY FACTORS

Regulatory Complexity and Institutional Diversity

Technical education in India operates within a multi-layered regulatory architecture involving central agencies, state authorities, professional councils, and accreditation bodies [2,5,6,18]. While regulation is essential to ensure minimum standards and public accountability, procedural complexity and regulatory rigidity can sometimes constrain institutional innovation and responsiveness to emerging educational needs.

Uniform regulatory norms applied across highly diverse institutions may not sufficiently account for variations in institutional scale, mission, regional context, and capacity [4–6]. A small rural engineering college and a nationally funded institute of technology often operate within broadly similar compliance frameworks despite vastly different capacities. Such standardization can unintentionally encourage a compliance-driven or "checklist" approach, in which meeting prescribed norms becomes the primary institutional objective rather than fostering continuous academic improvement, institutional distinctiveness, and innovation.

Evaluation Metrics and Academic Incentives

An increasing emphasis on research funding, publication metrics, and institutional rankings—reflecting broader global trends—has reshaped academic incentive structures [7,8]. When evaluation systems disproportionately reward research income, publication counts, or ranking indicators, excellence in teaching, curriculum development, and student mentoring may receive comparatively less recognition. Balanced academic ecosystems therefore require evaluation frameworks that recognize excellence in teaching, research productivity, innovation, and industry engagement [4,7,8].

Financial Sustainability and Institutional Stability

Financial architecture constitutes another critical structural determinant of educational quality [4–6]. Public institutions depend substantially on government grants, which may vary according to fiscal priorities and policy shifts. Private institutions rely largely on tuition revenue, making them sensitive to fluctuations in enrolment and seat occupancy.

During periods of declining admissions, financial pressures may influence faculty recruitment, research investment, laboratory modernization, and infrastructure maintenance [4,5]. Sustainable educational quality therefore requires funding models that provide stability and predictability while encouraging institutional performance, accountability, and continuous quality improvement.

Faculty Development and Capacity Building

Opportunities for research collaboration, industry immersion, pedagogical training, and exposure to emerging technologies remain unevenly distributed across institutions [2,7,8]. As these activities are essential for maintaining academic quality and curricular relevance, faculty members must continually update both their disciplinary expertise and pedagogical skills. Without systematic faculty upskilling, continuous professional development, and institutional support for adopting emerging technologies, curricular reform alone cannot ensure academic relevance or meaningful academia–industry alignment.

Reform Initiatives and Implementation Challenges

Recent policy initiatives—including those articulated in the National Education Policy 2020—have emphasized multidisciplinary education, academic autonomy, credit mobility, and outcome-based frameworks [2]. However, policy intent does not automatically

translate into academic transformation. Implementation capacity varies considerably across institutions [4–6]. Structural reform must therefore be accompanied by targeted capacity-building, governance reform, sustained institutional support, and periodic evaluation of outcomes to achieve lasting improvements in educational quality and academia–industry alignment.

THE WAY FORWARD: FROM MASSIFICATION TO MEANINGFUL ALIGNMENT

Having achieved substantial expansion in access, Indian technical education must now transition from massification to meaningful alignment through consolidation, qualitative improvement, and strengthened institutional capacity. The objective is not to reverse expansion but to align future growth with institutional capacity, educational quality, academic depth, and labour market relevance. The challenges discussed in the preceding sections indicate that future reforms should focus less on quantitative expansion and more on enhancing the quality, relevance, resilience, and long-term sustainability of technical education [2,4,8]. The next stage of development must therefore emphasize institutional differentiation, academic excellence, innovation, and sustainable growth.

Institutional Differentiation and Functional Complementarity

A differentiated institutional ecosystem, based on functional complementarity rather than rigid hierarchy, can enhance systemic efficiency and improve resource utilization [3,19]. Not all institutions are expected to perform identical roles, nor should they be evaluated against uniform benchmarks. A differentiated institutional framework may therefore be envisaged.

i. Research-intensive institutions, pursuing advanced research, postgraduate education, innovation, and externally funded projects while maintaining excellence in teaching.

ii. Regionally anchored institutions, including engineering colleges affiliated to state universities and well-equipped private institutions, emphasizing strong laboratory infrastructure, applied research, and close industry integration. Structured industrial exposure, including semester-long internships or vocational engagement, could help students develop familiarity with real-world technical challenges. Strong regional linkages with local manufacturing and service industries can further enhance relevance and graduate employability.

iii. Skill- and practice-oriented institutions, including polytechnics and Industrial Training Institutes, prioritizing intensive hands-on training, technical competence, and employability. Strengthened laboratory engagement and apprenticeship models can build student confidence in executing assigned tasks while fostering problem-solving ability, innovation, and workplace readiness.

Such differentiation acknowledges varied student aspirations and labour market needs while maintaining national coherence [3,19]. Clear mission statements, aligned evaluation criteria, and performance-linked funding mechanisms may reduce unrealistic uniformity and compliance-driven functioning while strengthening graduate employability and institutional responsiveness to evolving workforce requirements [4–6,17,20].

In skill- and practice-oriented institutions, systematic integration of laboratory learning, industrial exposure, and apprenticeships remains essential. Effective apprenticeship pathways can significantly improve employment prospects and entrepreneurial readiness [17,20].

Deepening Industry Engagement

Industry engagement must evolve from sporadic interactions to sustained institutional partnerships. Such partnerships should be long-term, strategic, outcome-oriented, and mutually beneficial, extending beyond student internships to encompass curriculum co-design, faculty development, consultancy, joint research, innovation, technology transfer, and continuing professional education [19]. Curriculum co-design committees, long-term internship agreements, joint research centres, technology transfer offices, adjunct industry faculty appointments, faculty immersion programmes, and collaborative innovation hubs can substantially strengthen academia–industry alignment [15,17,20]. Such sustained academia–industry collaboration can improve curriculum relevance,

graduate employability, research translation, innovation, and the overall responsiveness of technical education to evolving industrial and societal needs.

Faculty Development and Academic Continuity

Faculty capacity remains central to meaningful reform. Continuous improvement in teaching also benefits from systematic mechanisms for formative student feedback, peer review, and reflective teaching practice, enabling faculty to refine instructional methods and enhance learning outcomes. Equally important, continuous professional development in emerging technologies, pedagogical innovation, research methodology, and industry practices is critical for sustaining instructional quality and curricular relevance [2,7,8].

Encouraging interdisciplinary collaboration, industry-supported projects, and applied research can strengthen the integration of theoretical instruction with real-world engineering practice.

In disciplines experiencing faculty shortages, the structured engagement of experienced retired faculty members and technical experts—where appropriate—can provide valuable mentorship, strengthen laboratory instruction, and ensure academic continuity [2,7,8].

Rationalization and Sustainable Capacity

In regions facing persistent under-enrolment, strategic rationalization of intake capacity may be necessary. Institutional consolidation, when carefully implemented, can prevent resource dilution and enable the remaining institutions to operate more efficiently and sustainably [4,5]. Sustainable educational quality requires stable enrolment, adequate faculty strength, modern infrastructure, and predictable funding—conditions that are difficult to achieve under chronic overcapacity [4–6].

Broadening the Research Base

While centres of excellence continue to perform at high levels, strengthening the research ecosystem across a broader institutional base remains important. Expanding access to competitive research funding, developing shared laboratory facilities, and fostering regional research clusters can gradually reduce the concentration of research output in a limited number of elite institutions.

Research-informed teaching enhances student engagement, innovation capability, and industry relevance. Collaborative research involving academia, industry, and government agencies through regional innovation clusters can further accelerate the translation of research into technological innovation, industrial application, and societal impact [6,7,19].

Shifting the Metric of Success

The emphasis must shift from quantitative indicators—such as the number of institutions, sanctioned seats, and enrolment figures—to qualitative measures including learning outcomes, research impact, industry collaboration, innovation output, and graduate employability and career progression.

Expansion was necessary; consolidation must now ensure depth [2,19].

International Lessons and Contextual Adaptation

Several countries have addressed the challenges of expansion through differentiated institutional missions and stronger integration between higher education and national development priorities. International experience demonstrates that scale alone does not guarantee excellence.

Germany's dual vocational–technical model emphasizes apprenticeships and structured industry participation. In the United States, research universities integrate undergraduate education with strong research ecosystems, while China's higher education expansion has been accompanied by targeted investment in research-intensive universities to strengthen national innovation capacity.

Together, these examples illustrate that differentiated institutional missions, strategic investment, and strong industry engagement are central to converting quantitative expansion into qualitative excellence. India's reform trajectory can therefore benefit from adapting these underlying principles to its own socio-economic context rather than replicating institutional models mechanically [3,8,19].

CONCLUSION

India has successfully expanded access to higher and technical education on an unprecedented scale,

representing one of the country's most significant public policy achievements in recent decades. The next phase of reform should ensure that this quantitative success is matched by sustained improvements in educational quality, institutional capacity, and meaningful academia–industry alignment [2,4].

However, expansion that is not accompanied by proportional strengthening of faculty capacity, research ecosystems, laboratory infrastructure, governance mechanisms, and sustained industry engagement inevitably leads to variations in institutional quality and graduate outcomes [3,4]. Such heterogeneity is neither unexpected nor inherently undesirable in a country of India's scale and diversity; similar structural tensions have accompanied massification in higher education systems worldwide. Nevertheless, if left unaddressed, these imbalances may constrain academic quality, innovation capacity, and graduate employability.

The present phase therefore calls not for further expansion alone, but for consolidation, differentiation, and qualitative strengthening. Clear institutional missions, stable funding mechanisms, systematic faculty development, modern laboratory infrastructure, curriculum renewal, meaningful academia–industry collaboration, and robust quality assurance together provide the foundation for sustainable improvement [2,19]. The emphasis must progressively shift from quantitative indicators of growth towards qualitative measures of learning outcomes, research relevance, innovation, graduate employability, and societal impact.

Indian technical education now stands at a critical transition—from expansion-driven massification towards institutional maturation and quality enhancement. The trajectory of the coming decade will determine whether impressive numerical growth is translated into sustained academic excellence capable of supporting innovation-led, technology-driven, and employment-oriented national development. Ultimately, the challenge is not one of scale alone, but of coherence: only when expansion is aligned with institutional capacity, educational quality, clarity of purpose, and responsiveness to societal and industrial needs can it provide a durable foundation for India's long-term technological competitiveness and inclusive economic development.

REFERENCES

1. Trow M. Reflections on the Transition from Elite to Mass to Universal Access: Forms and Phases of Higher Education in Modern Societies. In: Forest JJ, Altbach PG (eds.), International Handbook of Higher Education. Springer, Dordrecht, 2007.
2. Government of India, National Education Policy 2020, Ministry of Education, New Delhi, 2020.
3. Altbach, P. G., Reisberg, L. and Rumbley, L. E., Trends in Global Higher Education: Tracking an Academic Revolution, UNESCO, Paris, 2009.
4. All India Survey on Higher Education (AISHE), All India Survey on Higher Education 2021–22, Ministry of Education, Government of India, New Delhi, 2023.
5. All India Council for Technical Education (AICTE), Approval Process Handbook 2023–24, AICTE, New Delhi, 2023.
6. University Grants Commission (UGC), Annual Report 2021–22, UGC, New Delhi, 2022.
7. Department of Science and Technology (DST), Research and Development Statistics 2022–23, Government of India, New Delhi, 2023.
8. OECD. Education at a Glance 2025: OECD Indicators. OECD Publishing, Paris, 2025. doi:10.1787/1c0d9c79-en.
9. Hao Y, Liu Y, Liu B, Amarantidis G, Ghannam R. Integrating AI in Engineering Education: A Comprehensive Review and Student-Informed Module Design for U.K. Students. IEEE Transactions on Education. 2025;68(2):173–185. doi:10.1109/TE.2025.3536105.
10. Rahman A, Khandakar A, Ayari MA, Naji K, Al Ali AK, Sellami AL, Alhazbi SMA. Artificial Intelligence Innovations, Challenges and Emerging Trends in Engineering Education. Discover Education. 2026;5, article 179. doi:10.1007/s44217-026-01137-1.
11. Chamola V, Dave D, Goyal I, Sharma S. Generative Artificial Intelligence in STEM Education: A Review of Applications, Benefits and Challenges. Expert Systems. 2026;43(2):e70201. doi:10.1111/exsy.70201.
12. Segura Altamirano SF, Maquen Niño GLE, Guzmán Roldán CM, Pérez Herrera A, Castro Cárdenas DM, et al. Generative Artificial Intelligence in Higher Education: A Systematic Review of Perceptions, Implementation

and Pedagogical Transformation. Review of Education. 2026;14:e70152. doi:10.1002/rev3.70152.

13. Bacalso FD, Villamor FE, Paclipan CP, Orquia CJB, Rendon IP, Albiso AT, Himang CM. Institutional Approaches to Generative AI Management in Higher Education: A Systematic Review. Frontiers in Education. 2026;11:1814426. doi:10.3389/feduc.2026.1814426

14. de Silva A, Song IHJ, Yang H, Humayoun SR. STEM Faculty Perspectives on Generative AI in Higher Education. Proceedings of the AAAI Symposium Series. 2026;8(1):640–647. doi:10.1609/aaaiss.v8i1.42600 (preprint: arXiv:2603.04001)

15. AI4SG Lab, NUS School of Computing, SoC Horizons Office, and Google. Reshaping Undergraduate Computer Science Education in the Generative AI Era (Version 1.0). White Paper, 2026. arXiv:2606.07545. doi:10.48550/arXiv.2606.07545.

16. Alenezi M. The Rise of AI Native Software Engineering: Implications for Practice, Education, and the Future Workforce. arXiv preprint, June 12, 2026. arXiv:2606.12986. doi:10.48550/arXiv.2606.12986.

17. World Economic Forum. The Future of Jobs Report 2025. World Economic Forum, Geneva, Switzerland, 2025.

18. National Board of Accreditation (NBA). Manual for Accreditation of Undergraduate Engineering Programs (Tier I). National Board of Accreditation, New Delhi, 2025.

19. UNESCO. Engineering for Sustainable Development: Delivering on the Sustainable Development Goals. UNESCO, Paris, 2021.

20. India Skills Report 2026. Wheebox, Confederation of Indian Industry (CII), AICTE, Association of Indian Universities (AIU), New Delhi, 2026.

DOI - 10.68648/ijte-q-2026-49-03-019

Mechanical Durability of Contemporary Dental Materials: An Engineering Review and Research Roadmap

Amol B. Patil
PhD Research Scholar
Department of Mechanical Engineering
Srinivas University
Mangalore, Karnataka
✉ amolbpatil111@gmail.com

Amit S. Shelake
Associate Professor
Departmennt Mechanical Engineering
Ashokrao Mane Group of Institutions
Kolhapur, Maharashtra
✉ shelakeamit111@gmail.com

Jitendra B. Patil
PhD Research Scholar
Department of Mechanical Engineering
Srinivas University
Mangalore, Karnataka
✉ jitendrapatil11691@gmail.com

ABSTRACT

Dental restorations and implants operate under coupled cyclic loading, contact, sliding, moisture, thermal fluctuation, saliva chemistry and manufacturing-induced defects; therefore, static strength alone is an inadequate predictor of service life. This critical review synthesizes 50 peer-reviewed studies published in 2025-2026 and interprets zirconia/glass-ceramics, resin composites and additively manufactured polymers, PEEK/high-performance polymers, and titanium/Co-Cr systems through a process-structure-property-environment-life framework. Evidence shows that build orientation, layer thickness, firing or post-curing, milling, surface treatment and ageing alter defect populations and fatigue reliability; printed-resin durability remains protocol-dependent; and implant performance requires simultaneous treatment of micromotion, stress redistribution, corrosion and tribocorrosion. The review integrates service-representative testing, fracture/fatigue and wear metrics, reliability analysis and validated finite-element modelling into a compact experimental-computational methodology. Priority directions are standardized fatigue/wear protocols, defect-informed additive-manufacturing models, calibrated ageing, coupled bio-tribocorrosion-fatigue testing, experimentally validated probabilistic FEA, and multi-objective material-process-geometry optimization.

KEYWORDS *: Dental materials; Fatigue; Tribology; Additive manufacturing; Finite element analysis.*

NOTATIONS

a - flaw/crack size; b - specimen width or fatigue exponent (context dependent); C, m - crack-growth constants; d - specimen thickness; F - applied/fracture load; f - loading frequency; H - hardness; K_I - mode-I stress intensity factor; L - support span or sliding distance; N, N_f - cycle count and cycles to failure; P_f - failure probability; R - load ratio; V - wear volume; Y - geometry factor; σ, σ_a - stress and stress amplitude; σ0 - Weibull characteristic strength; FEA - finite-element analysis; DOE - design of experiments; PEEK - polyetheretherketone; SEM - scanning electron microscopy; XRD - X-ray diffraction; micro-CT - micro-computed tomography; DIC - digital image correlation.

INTRODUCTION

Contemporary restorative dentistry increasingly relies on materials and manufacturing routes that are also central to mechanical and materials engineering: transformation-toughened ceramics, glass-ceramics, particle- and fiber-reinforced polymers, high-performance thermoplastics, titanium alloys, cobalt-

chromium alloys, CAD/CAM machining and additive manufacturing. What differentiates a dental component from a conventional engineering component is not the absence of mechanics, but the unusually coupled service environment. A crown, veneer, denture base, abutment or implant experiences repeated multiaxial loading while its surfaces are simultaneously exposed to moisture, temperature excursions, pH changes, saliva, abrasive food particles, antagonistic contact and-in metallic systems-electrochemical reactions. The engineering question is therefore not merely whether a material is 'strong', but how defects initiate, evolve and interact until function is lost.

The 2025-2026 literature reinforces this distinction between strength and durability. Fatigue studies on lithium disilicate, multilayer zirconia and monolithic crowns show that thickness, contact configuration, support conditions, surface treatment and loading protocol change failure probability even when the nominal material class is unchanged [1, 2, 3, 4, 5]. In resin systems, print orientation, polymerization/post-curing and artificial ageing alter flexural response and fracture resistance, while clinical and laboratory wear studies demonstrate that surface damage develops progressively under function [6, 7, 8, 9]. These findings support a damage-mechanics interpretation in which processing determines the defect state, the defect state controls local stress amplification, and the service environment controls damage accumulation.

Digital manufacturing intensifies the need for this engineering viewpoint because it introduces controllable process variables and process-specific defects. Milled components may contain machining grooves, edge chipping, residual stresses or subsurface damage; vat-polymerized resins may exhibit orientation-dependent layer interfaces, incomplete conversion or post-cure gradients; printed zirconia may contain pores, interlayer defects and shrinkage-related heterogeneity; and printed metals may retain porosity, surface roughness and microstructural gradients. Systematic evidence published in 2025-2026 generally indicates that printed zirconia can achieve clinically relevant strength but still tends to show lower flexural strength or Weibull reliability than milled zirconia, with build orientation, layer thickness and firing history remaining important variables [10, 11, 12]. Bergler et al. [13] further connected flexural response with microstructural differences between 3D-printed and milled mono- and multilayer zirconia, reinforcing the need to characterize process-induced defects rather than report strength alone. Likewise, recent reviews of definitive printed resins emphasize heterogeneity and unresolved long-term durability relative to CAD/CAM-milled counterparts [14, 15].

A Mechanical Engineering PhD can contribute uniquely to this field by combining experimental mechanics, tribology, fracture and fatigue, surface engineering, metrology, additive manufacturing, finite-element analysis (FEA), design of experiments (DOE), reliability analysis and optimization. Such a contribution is most valuable when it goes beyond a simple comparison of commercial materials and establishes a causal chain: processing variable -> microstructure/defect -> local stress or contact state -> damage mechanism -> life/reliability -> design rule. Dynamic FEA studies already show that framework and abutment stiffness redistribute stress across implant components, but numerical predictions require experimental validation and realistic contact/loading assumptions before they can become predictive design tools [16, 17].

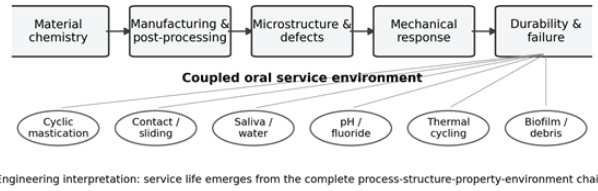


Fig. 1. Mechanical-engineering framework linking chemistry, manufacturing, defects, mechanical response and durability in the oral environment

This review therefore has two objectives. First, it synthesizes recent 2025-2026 evidence on mechanical behaviour, damage mechanisms and durability across the principal contemporary dental material families. Second, it converts that synthesis into a research roadmap for Mechanical Engineering scholars by identifying measurable variables, experimental methods, modelling strategies, knowledge gaps and candidate doctoral objectives. The central thesis is that dental durability should be treated as a coupled process-structure-property-environment-life problem rather

than as a collection of isolated strength values. Fig. 1 summarizes this organizing framework.

LITERATURE REVIEW / SURVEY

Zirconia and glass-ceramics

The contemporary ceramic landscape is shifting from a simple zirconia-versus-lithium-disilicate comparison toward composition gradients, ultra-translucent zirconias, multilayer blanks, minimally invasive thicknesses and additive manufacturing. These innovations improve esthetics and manufacturing flexibility but increase mechanical complexity. Multilayer zirconia is spatially non-uniform; a restoration's position within the blank can alter phase/composition and thus stiffness, strength or fatigue behavior. Bakitian [18] synthesizes manufacturer, laboratory, clinical and ageing factors affecting multilayer zirconia, making a strong case for layer-resolved reporting rather than treating a multilayer disc as homogeneous.

Fatigue and thickness studies reveal that material ranking is geometry dependent. Zurlo Dellazzana et al. [4] found that thickness affected 4YSZ and lithium disilicate differently in occlusal-veneer fatigue. Prott et al. [5] reported distinct fatigue/failure behavior for lithium disilicate and composition-gradient zirconia veneers across 0.5-1.5 mm thicknesses. Such data are ideal for developing a mechanistic 'minimum safe thickness' framework in which the relevant output is failure probability under a defined load spectrum rather than a universal thickness number.

Surface engineering can improve bonding, bioactivity or antimicrobial function but may also create stress raisers or interface failure modes. Cantelli et al. [2] and da Silva et al. [19] demonstrate that surface/coating interventions need fatigue evaluation and emphasize the dual response of crown fatigue and antagonist wear. The Mechanical Engineering opportunity is to optimize surface treatments using fatigue life, coating adhesion, roughness and antagonist loss simultaneously.

Additive zirconia remains a particularly fertile PhD topic. Pinelli et al. [10] and Ruggiero et al. [11] identify lower or more variable flexural reliability of printed material relative to milled controls and emphasize orientation/layer effects. Sabatini et al. [12] directly links firing process and build orientation with flexural/structural response, whereas Bergler et al. [13] adds microstructural comparison across printed and milled mono- and multilayer zirconia. A compelling doctoral program would combine print DOE, uCT porosity mapping, XRD phase analysis, surface topography, flexure/Weibull, step-stress fatigue and fracture-surface statistics, followed by a defect-informed finite-element or probabilistic life model.

Resin composites and 3D-printed polymers

Resin-based dental materials span direct composites, CAD/CAM polymer blocks, denture-base polymers, short-fiber systems and printable resins. Their durability is governed by a network of coupled mechanisms: polymerization shrinkage and residual stress, matrix viscoelasticity, filler-matrix debonding, water uptake, hydrolysis, wear and fatigue. Ribeiro et al. [20] illustrates how curing technology affects physicomechanical properties, while Mohammadi Ghayeghchi et al. [21] demonstrates chemistry-driven control of polymerization shrinkage stress. These mechanisms are directly accessible to Mechanical Engineering through DMA, shrinkage-force measurement, micro/nanoindentation, fracture testing and FE residual-stress analysis.

The strongest recent theme is additive manufacturing. Coldea et al. [22], Artopoulou et al. [23] and Alhotan et al. [24] show that denture-base mechanics cannot be generalized across printed formulations. Rizzante et al. [6] and Korkmaz et al. [9] show the importance of polymerization/post-curing protocols; Mudhaffer et al. [7] adds build orientation and ageing. Terán Lozada et al. [25] evaluates definitive crown resins, while Benalcázar-Jalkh et al. [26] brings fatigue into implant-supported definitive crowns. The umbrella synthesis by Dascanio et al. [15] concludes that long-term durability of printed resin-based composites remains uncertain and that milled CAD/CAM materials generally retain stronger mechanical/optical stability. This is precisely the type of evidence gap a Mechanical Engineering PhD can address with standardized fatigue and damage characterization.

Wear studies extend beyond static properties. Doumit et al. [8] provides 24-month functional wear data for 3D-printed minimally invasive restorations, creating an important bridge between laboratory simulation and

clinical behavior. Ghabchi et al. [27], Li et al. [28], Wu et al. [29] and Osiewicz et al. [30] demonstrate that material loss depends on both material architecture and tribological configuration. A high-value research design would calibrate a laboratory wear protocol against clinically observed roughness/volume-loss evolution rather than simply increasing cycle count.

Direct composites remain relevant because they reveal the interaction between intrinsic material design and restoration geometry. Tsujimoto et al. [31] found that material/filler type and the number of restored surfaces were associated with restoration longevity in a large retrospective dataset. Troha et al. [32] shows that filler content controls ageing response, and Watanabe et al. [33] addresses short-fiber reinforcement under fatigue. These studies suggest an integrated research direction: use clinical failure/survival patterns to select realistic geometries and loads, then resolve their mechanics experimentally and numerically.

PEEK and high-performance polymers

PEEK is attractive because its low modulus relative to ceramics/metals can change load transfer and because filler modifications can alter wear behavior. It should not be evaluated using strength alone. Li et al. [28] showed that certain filled CAD/CAM PEEK compounds achieved favorable simulated-chewing wear and low antagonist abrasiveness, with plastic deformation and debris-layer effects contributing to the mechanism. Sari et al. [34] compared PEEK and zirconia implant abutments produced by subtractive and additive routes after thermomechanical ageing, demonstrating the importance of manufacturing route and connection behavior.

PEEK also exposes a key modelling issue: a compliant component can reduce stress locally while increasing displacement or redistributing stress elsewhere. Dynamic and hybrid-abutment FEA studies show that material modulus changes the stress field across the framework/Ti-base/bone system [16, 17]. Therefore, a 'lower stress in the PEEK component' is not automatically a better biomechanical outcome. Research must evaluate the whole assembly, micromotion/contact, fatigue and bone-level response with experimental validation. Ateş and Gungormus [35] further show that thermal ageing responses differ among PEEK, fiber-reinforced and graphene-reinforced CAD/CAM polymers, reinforcing the need for material-specific constitutive data.

Titanium, Co-Cr and implant systems

Metallic dental systems generally possess greater fracture toughness than ceramics, but their durability is governed by different weak links: notch-sensitive fatigue at threads and connections, preload loss, fretting, corrosion and tribocorrosion. Schettini et al. [36] shows that Co-Cr microstructure and mechanical properties depend on whether the alloy is cast, milled or printed. For titanium, Vegas-Bustamante et al. [37], Mangal et al. [38], Hariprasad et al. [39] and Boffano et al. [40] collectively shift the focus from passive corrosion toward coupled mechanical-electrochemical degradation.

Table 1. Engineering taxonomy of contemporary dental material systems.

Material system	Dominant durability concern	High-value research direction
Zirconia / multilayer ceramics	Brittle fracture, slow crack growth, AM/machining flaws, ageing	Thickness/gradient design; AM defect control; fatigue-surface interaction
Lithium disilicate glass-ceramic	Surface flaws, contact fracture, moisture-assisted crack growth	Damage-tolerant minimum thickness; support/contact effects
Resin composites	Matrix fatigue, filler debonding, shrinkage, wear, hydrolysis	Filler-interface ageing; repair fatigue; process-life modelling
3D-printed photopolymers	Anisotropy, conversion gradients, water sorption, fatigue uncertainty	Orientation/post-cure optimization; long-term ageing calibration
PEEK / high-performance polymers	Creep, wear, bonding and interface failure	Fatigue-wear; filler/self-lubrication; surface activation
Ti / Ti alloys / Co-Cr	Fatigue, fretting, corrosion, tribocorrosion, ion release	Coupled electrochemical-mechanical life; AM/surface engineering
Interfaces / adhesives	Debonding, hydrolysis, mixed-mode cyclic damage	Interface fracture mechanics and life prediction

Alloy and surface design provide a materials-engineering route to improvement. Herbster et al. [41] investigates Cr additions to titanium implant alloys with the objective of improving corrosion/wear response. Yet any alloying or coating strategy must preserve biocompatibility and fatigue resistance; maximizing corrosion resistance while reducing fatigue strength would not constitute a successful dental material. Multi-objective optimization using corrosion current, wear, fatigue life, hardness, modulus and biological constraints is therefore more defensible than a single-property optimization.

For implant components, component-level testing and FEA should converge. Padullés-Roig et al. [42] demonstrates this combination after implantoplasty, while Sahin Hazir et al. [16] and Poovarodom et al. [17] show the sensitivity of stress distribution to framework and abutment materials. The research frontier is experimentally validated dynamic/contact FEA that includes realistic preload, nonlinear interfaces, patient-specific geometry, cyclic load spectra and material degradation.

METHODOLOGY: REVIEW DESIGN AND EVIDENCE FRAMEWORK

This article is a structured critical review rather than a formal meta-analysis. A focused literature search was updated through August 2026 using PubMed/MEDLINE and publisher-indexed web searches. Search combinations included dental material, zirconia, lithium disilicate, resin composite, 3D-printed resin, denture base, PEEK, titanium, cobalt-chromium, implant, fatigue, fracture, wear, tribology, corrosion, tribocorrosion, thermomechanical ageing, CAD/CAM, additive manufacturing and finite element. Priority was given to peer-reviewed papers with publication years 2025 or 2026 that reported mechanical, durability, manufacturing or modelling outcomes directly relevant to a Mechanical Engineering research question.

Fifty papers were selected as a deliberately recent evidence corpus: 25 from 2025 and 25 from 2026. The corpus includes primary laboratory studies, clinical/mechanical studies, systematic reviews, an umbrella review and computational investigations. It is not intended to replace broader historical literature; instead, it captures the current problem frontier and is used to identify where mature test concepts coexist with unresolved standardization and translation issues. Fig. 4 quantifies the corpus by year and primary material domain.

Evidence was synthesized along five engineering axes: (i) material system and microstructure; (ii) manufacturing and post-processing history; (iii) loading/contact/environmental boundary conditions; (iv) measured response or failure mode; and (v) unresolved mechanism or validation need. This framework avoids the common problem of comparing strength values obtained under incompatible geometries, supports, loading rates, ageing histories or failure definitions. A result is considered most useful for doctoral research when it reveals a controllable engineering variable and a mechanistic response, not simply a statistically significant material ranking.

RESULTS AND DISCUSSION

Oral service environment, contact and cumulative damage

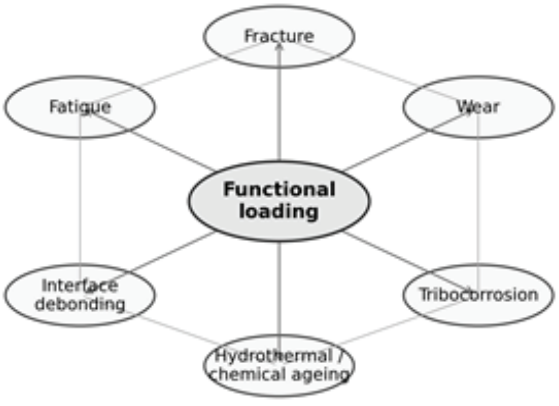
The oral cavity imposes a variable-amplitude, multiaxial and contact-dominated loading history. Mastication repeatedly loads restorations through a compliant chain comprising enamel/dentin or implant support, luting layer, restorative material and antagonist. Contact location migrates, friction introduces tangential traction, and local geometric features-fissures, connectors, screw-access channels, finish lines and layer transitions-produce stress concentrations. Therefore, a single axial compression test can rank materials but cannot reproduce the dominant damage mechanisms of many clinical components. The sensitivity of ceramic crowns to contact pattern and core material reported by Packaeser et al. [3], and the thickness-dependent fatigue behaviour reported for occlusal veneers by Zurlo Dellazzana et al. [4] and Prott et al. [5], illustrate why geometry, support and contact must be treated as test variables rather than incidental details.

Cyclic load magnitude is only one element of the spectrum. Frequency controls test time and can influence viscoelastic or time-dependent materials; waveform changes dwell time; load ratio changes mean stress; sliding amplitude determines whether the problem resembles indentation fatigue, two-body wear or fretting; and antagonist geometry controls contact pressure. In fatigue experiments the relevant descriptors include maximum and minimum load, stress amplitude, mean stress, load ratio R, frequency f, number of cycles N, failure definition and run-out criterion. When a step-stress protocol is used, the increment schedule is part of the mechanical model and should not be compared directly with constant-amplitude data without appropriate statistics [1, 4].

Temperature and chemistry modulate these mechanics. Thermocycling generates repeated differential

expansion between matrix and filler, core and veneer, resin and ceramic, or implant and abutment. Water can plasticize polymer matrices, facilitate hydrolysis at silane-treated filler interfaces and change crack-growth kinetics. pH, fluoride and electrochemical potentials can destabilize passive titanium oxide, while micromotion repeatedly removes or disrupts the film. The resulting tribocorrosion is not the sum of wear and corrosion measured independently; the mechanical and electrochemical processes interact. Recent titanium evidence continues to emphasize this coupling and the need for service-representative protocols [37, 38, 39, 40].

For polymers and composites, ageing should be considered as a state variable. Thermal cycling, prolonged water storage and solvent exposure may alter modulus, hardness, residual stress and interface strength before cyclic loading is applied. Troha et al. [32] reported that filler amount and ageing protocol jointly influence the long-term mechanical stability of experimental composites, while Ateş and Gungormus [35] observed material-dependent surface and mechanical changes after thermal ageing of CAD/CAM polymers. These studies highlight a key methodological point: 'ageing' is not a universal dose. Its severity depends on material chemistry, exposure duration, thermal range, solution and whether mechanical loading is applied concurrently or sequentially.



Damage modes interact; ageing changes interfaces, wear changes contact stress, and corrosion modified fatigue initiation

Fig. 2. Coupled damage mechanisms: fatigue, fracture, wear, ageing, interface degradation and tribocorrosion can accelerate one another.

Table 2. Service variables and laboratory representation for durability research

Service feature	Laboratory representation / response	Critical reporting / interpretation
Repeated mastication	Servo/electrodynamic fatigue or chewing simulation; life, survival, stiffness loss	Fmax/Fmin, R, f, N, waveform, run-out and failure criterion
Contact / sliding	Reciprocating/ball-on-disc/chewing simulation; wear depth/volume, COF	Antagonist material/radius, pressure, path, lubricant and distance
Thermal / moisture ageing	Thermocycling and water/artificial saliva storage; property retention	Temperature range, dwell, chemistry and exposure time
Acid / fluoride / biofilm	Immersion/electrochemistry/biofilm-informed media; current, ion release	pH, ions, potential and simultaneous mechanical contact
Implant micromotion	Cyclic implant-abutment loading/fretting; preload loss, microgap, fatigue	Preload, displacement amplitude, friction/contact and geometry
Manufacturing defects	micro-CT/SEM/profilometry plus mechanical test; reliability/crack origin	Pore/flaw size, orientation, roughness and process history

A useful Mechanical Engineering interpretation of mastication is a moving contact problem superimposed on a structurally compliant assembly. The antagonist generates a local normal pressure field and, whenever sliding occurs, a tangential traction field. The resulting tensile stress that governs brittle fracture may occur away from the point of maximum compressive contact pressure, particularly near the intaglio surface, margins, connectors or interfaces. This distinction explains why compressive strength is a weak predictor of crown or veneer survival. A mechanically meaningful experiment should therefore document indenter/antagonist radius, contact location, support stiffness, sliding path and frictional condition in addition to the applied force. Packaeser et al. [3], Zurlo Dellazzana et al. [4], Prott et al. [5] and Spitznagel et al. [43] collectively reinforce the sensitivity of fatigue performance to contact pattern, restoration design and thickness.

For ceramic and hybrid restorations, three stress scales should be distinguished. The first is the nominal structural stress governed by thickness, curvature and support. The second is the contact-scale stress beneath the antagonist, which depends on local curvature and elastic mismatch. The third is the defect-scale stress amplification at pores, machining grooves, interlayer boundaries or edge flaws. Failure occurs when these scales interact so that a critical flaw experiences a crack-driving force above the material's resistance. This hierarchy suggests a powerful PhD strategy: characterize defect population experimentally, calculate structural/

contact stress numerically, and connect the two through probabilistic fracture or fatigue. The additive-zirconia comparisons of Pinelli et al. [10], Ruggiero et al. [11], Sabatini et al. [12] and Bergler et al. [13] are especially suitable for this process-defect-stress linkage.

Polymer-based restorations require an additional time-dependent layer. Their contact area and stress field can evolve because modulus, creep and viscoelastic relaxation depend on temperature, water uptake and loading rate. Thus, the same nominal load can produce a different local deformation after ageing. For printed resins, build orientation and post-curing may also make the response anisotropic or spatially non-uniform [7, 6, 9]. A robust research design should therefore pair contact or component testing with orientation-resolved modulus and strength, and should re-measure key properties after environmental conditioning rather than using as-printed values in FEA.

Implant-supported systems add screw preload, interface contact, microgap and bone/support stiffness to the problem. The load path is redistributed through crown, abutment, screw, implant and surrounding support. Simplifying every interface as perfectly bonded suppresses micromotion and can artificially reduce local stress gradients. Dynamic FEA by Sahin Hazir et al. [16] and hybrid-abutment modelling by Poovarodom et al. [17] illustrate the importance of material stiffness and component geometry, while Padullés-Roig et al. [42] demonstrates the value of combining component fatigue with simulation. The next step is experimentally measured contact/micromotion and preload-sensitive model validation.

From load spectrum to cumulative damage

Dental loading is variable-amplitude rather than a single repeated force. From an engineering standpoint, the central research problem is how a sequence of subcritical events accumulates damage. In brittle ceramics, repeated contacts can extend pre-existing flaws by environmentally assisted slow crack growth. In polymers and composites, cyclic deformation can generate matrix microcracking, filler-interface debonding, fiber pull-out or viscoelastic damage. In metals, fretting marks, corrosion pits and thread-root stress concentrations can become fatigue origins. Accordingly, the same number of cycles is not equivalent when maximum load, mean load, sliding distance, environment or damage state differs.

A PhD study should distinguish an exposure descriptor from a damage descriptor. Exposure descriptors include cycles, temperature, pH, load, sliding distance and time. Damage descriptors include stiffness loss, crack length, wear depth, roughness, interfacial fracture energy, corrosion current, microgap or residual strength. The latter are more transferable because they describe the physical state of the specimen. For example, a thermocycling protocol becomes more scientifically useful when it is linked to a measurable modulus or bond-strength change, and a wear protocol becomes more useful when residual fatigue strength is measured after a defined wear depth. Troha et al. [32], Stape et al. [44], Wu et al. [29] and Osiewicz et al. [30] collectively motivate this transition from cycle counting to state-variable-based degradation.

This concept also avoids the common but weak practice of assigning an arbitrary number of laboratory cycles to a fixed number of clinical years. The proposed alternative is an equivalent-damage framework: choose one or more measurable damage variables, quantify their evolution under accelerated and less-accelerated exposures, and calibrate the laboratory severity against clinical or long-term observations where available. The 24-month clinical wear evidence reported by Doumit et al. [8] is particularly valuable as a potential calibration target for laboratory wear simulations.

Process-structure-property-durability relationships

The most transferable framework for dental-material research is the process-structure-property-durability chain. ‘Process’ includes powder or resin formulation, milling toolpath, printing orientation, layer thickness, exposure dose, debinding, sintering, firing, post-curing, heat treatment, polishing, air abrasion, coating and cementation. These operations create a ‘structure’ that includes grains, phases, interfaces, filler distribution, porosity, layer boundaries, residual stress and surface topography. The structure determines elastic modulus, strength, toughness, hardness, viscoelasticity and chemical stability; these properties then determine how damage accumulates under oral loading.

Zirconia is an especially clear example. Its measured strength is probabilistic because failure is flaw-

controlled; manufacturing determines the flaw population. Printed zirconia introduces additional sensitivity to build orientation, layer formation, debinding and sintering, while milling can introduce surface and subsurface damage. Recent systematic analyses conclude that milled zirconia generally retains higher flexural strength and Weibull reliability, although printed zirconia has substantial potential if process-induced defects are controlled [10, 11]. Sabatini et al. [12] further demonstrate that firing and build orientation are not secondary details but variables that alter structural and flexural behaviour. Wu et al. [45] broaden the comparison by coupling mechanical characterization with biosafety, an important reminder that mechanical optimization cannot be isolated from material compatibility.

Photopolymerized resins show a parallel but chemically different process dependence. Printing produces green parts whose final network is established by post-curing. Orientation changes the relationship between layer interfaces and principal stress; exposure and post-cure influence degree of conversion; washing can remove unreacted material but also affect surface state; and water uptake subsequently changes the polymer network. Rizzante et al. [6], Mudhaffer et al. [7] and Korkmaz et al. [9] collectively support the need to report the complete print-to-test history. Reviews published in 2026 identify protocol heterogeneity as a major limitation in comparing definitive printed resins and in claiming long-term equivalence with milled materials [14, 15].

Conventional resin composites demonstrate that formulation itself is a mechanical design variable. Filler amount, filler morphology, matrix chemistry, photopolymerization and interfacial silanization determine modulus, shrinkage stress, wear and hydrolytic stability. Mohammadi Ghayeghchi et al. [21] showed how addition-fragmentation chain-transfer chemistry can reduce polymerization shrinkage stress while maintaining acceptable flexural properties over an optimized concentration range. Troha et al. [32] showed that the benefit of higher filler loading depends on ageing history. These examples argue for multi-response optimization rather than maximizing a single property.

For metallic implants, manufacturing and surface state interact with electrochemistry. Additive manufacturing can change microstructure and roughness; machining and implantoplasty alter cross-section and surface condition; alloying changes passive-film and wear response. Schettini et al. [36] compared Co-Cr produced by casting, milling and 3D printing, while Herbster et al. [41] investigated Cr-containing titanium alloy concepts for corrosion and wear resistance. Padullés-Roig et al. [42] combined implantoplasty-related fatigue testing with finite-element simulation, illustrating a model PhD pattern: introduce a clinically relevant geometry/process change, measure its structural consequence, and validate a mechanics model.

Static properties, reliability and failure statistics

Static properties remain essential inputs, but their role should be interpreted mechanistically. Elastic modulus controls stiffness and stress transfer; hardness contributes to indentation/contact response but is not a direct substitute for wear resistance; flexural strength provides a convenient flaw-sensitive screening measure; fracture toughness quantifies resistance to crack extension; and yield/ultimate strength are relevant for ductile metallic components. For polymeric systems, storage modulus, loss factor, creep and stress relaxation may be as important as nominal flexural strength because time, temperature and moisture change the constitutive response.

For a rectangular specimen in three-point bending, the nominal flexural strength is commonly represented as:

$$\sigma_f = 3 F L / (2 b d^2) \quad (1)$$

where F is the fracture load, L the support span, b specimen width and d specimen thickness. The simplicity of this expression can mask sensitivity to edge finish, span-to-depth ratio, surface location under tension, loading rate and specimen defects. Consequently, strength data should be paired with fractography and, for brittle materials, reliability statistics. A two-parameter Weibull representation is useful when the goal is to quantify scatter rather than merely report a mean:

$$P_f = 1 - \exp[-(\sigma / \sigma_0)^m] \quad (2)$$

where m is the Weibull modulus and σ0 a characteristic strength. The 2025 systematic review by Pinelli et al.

[10] specifically highlights flexural strength and Weibull modulus as complementary metrics for printed versus milled zirconia. For doctoral work, Weibull analysis is most informative when specimen volume, stressed area, flaw origin and processing history are controlled, because otherwise m conflates true material variability with test inconsistency.

A second common misconception is to infer wear resistance directly from hardness. Recent PEEK and composite studies show why this is unsafe. Li et al. [28] found that filled PEEK compounds could display favorable impact-sliding wear despite lower baseline hardness than enamel, with deformation and debris-layer mechanisms contributing to performance. Osiewicz et al. [30] found that filler volume and hardness helped explain three-body wear in BPA-free composites but did not fully determine it. Wear is therefore an emergent tribosystem property defined by both surfaces, the intermediate medium, contact stress, motion and evolving debris.

Table 3. Compact mechanics and durability metrics

Metric / expression	Engineering use	Complementary method
Flexural: $\sigma_f = 3FL/(2bd^2)$	Flaw-sensitive static screening	Fractography + Weibull statistics
Weibull: $P_f = 1-\exp[-(\sigma/\sigma_0)^m]$	Strength scatter and reliability	Crack-origin mapping; stressed-area control
Fracture: $K_I = Y\sigma\sqrt{(\pi a)}$	Crack-driving force / defect sensitivity	Toughness / slow-crack-growth testing
Fatigue: $\sigma_a = \sigma_f (2N_f)^b$	Stress-life representation	Step-stress/S-N – survival statistics
Crack growth: $da/dN = C(\Delta K)^m$	Initial flaw to remaining life	Environment-specific calibration
Wear: $V = kWL/H$	First-order tribological descriptor	3D profilometry + contact analysis
Tribocorrosion: $V_{tot} = V_{mech} + V_{corr} + V_{syn}$	Synergistic electro-mechanical loss	Force/friction/electrochemistry

Fracture and fatigue mechanics

Ceramics fail when a pre-existing or process-generated flaw experiences a local stress intensity sufficient to extend. In linear elastic fracture mechanics, mode-I crack driving force is approximated by:

$$K_I = Y \sigma \sqrt{(\pi a)} \quad (3)$$

where a is flaw size and Y a geometry factor. This simple relation is valuable for dental research because it explains why two specimens of the same chemistry can show different strength after grinding, polishing, printing or coating: the process changes a, Y and residual stress. In zirconia, transformation toughening and phase composition further modify crack resistance; in glass-ceramics, subcritical crack growth in moisture can reduce long-term reliability. Surface-treatment studies such as Cantelli et al. [2] demonstrate that procedures intended to improve bonding also influence fatigue survival, so bond optimization must include the possibility of mechanically introduced flaws.

Restoration thickness changes both nominal stress and the crack-driving field. The 2026 occlusal-veneer study by Zurlo Dellazzana et al. [4] showed a material-thickness interaction: 4YSZ and lithium disilicate did not respond identically as thickness decreased. Prott et al. [5] similarly demonstrated that conclusions about ultrathin lithium disilicate cannot simply be transferred to composition-gradient multilayer zirconia. Spitznagel et al. [43] likewise showed that restoration configuration and thickness interact with fatigue performance and post-fatigue failure load in composition-gradient multilayer zirconia. For Mechanical Engineering, this creates a clear research opportunity for defect-sensitive thickness design using cohesive fracture, XFEM, probabilistic FEA or critical-defect mapping validated by fractography.

Fatigue is the central durability mechanism for many restorations because mastication applies millions of subcritical load cycles. In a stress-life representation, the Basquin-type relation provides a useful conceptual form:

$$\sigma_a = \sigma_f' (2N_f)^b \quad (4)$$

where σa is stress amplitude, σ'f a fatigue-strength coefficient and b a fatigue exponent. Dental literature often uses step-stress or chewing-simulator survival rather than complete S-N characterization, but the engineering principle remains the same: cyclic life depends on load amplitude, mean stress, environment, defect state and geometry. Fig. 3 illustrates why a high static-strength ceramic may still show substantial fatigue degradation while a lower-modulus polymer may dissipate or redistribute contact stresses differently. The curves are schematic and are not pooled data.

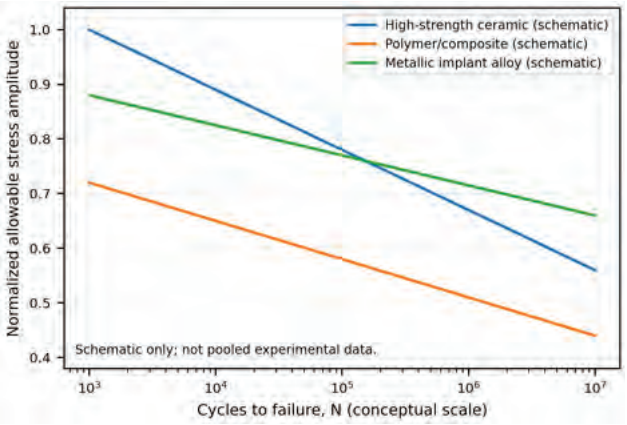


Fig. 3. Conceptual normalized S-N map across broad material classes; curves are schematic rather than pooled experimental data.

Recent fatigue studies offer several mechanistic lessons. Buser et al. [1] emphasize protocol sensitivity when comparing pressed lithium disilicate and 5Y-TZP. Packaeser et al. [3] show that core and occlusal-contact conditions alter ceramic-crown fatigue, while Pilecco et al. [46] links milling protocols to CAD/CAM fatigue. Benalcázar-Jalkh et al. [26] directly compares implant-supported 3D-printed and milled definitive resin crowns under fatigue. Watanabe et al. [33] brings reinforcement architecture into the problem by studying short-fiber flowable composite. Together these studies suggest that a high-value PhD should not ask 'which material is strongest?' but 'which process/geometry/environment combinations control fatigue initiation and propagation, and can the resulting life be predicted?'

Karevan et al. [47] further compared the fatigue behaviour of CAD/CAM composites with lithium-disilicate glass-ceramic, reinforcing the need for common cyclic-damage criteria across material classes.

Where crack-growth measurements are feasible, the Paris-type relation provides a mechanistic bridge between initial flaws and remaining life:

$$da/dN = C\,(\Delta K)^m \qquad (5)$$

The parameters C and m are environment- and material-dependent and should not be assumed from engineering handbooks for dental systems. For ceramics and bonded interfaces, subcritical growth may be strongly affected by moisture; for polymer composites, crack paths may deflect through matrix, filler interface or fiber bridges; and for metallic implants, corrosion-assisted fatigue can change both crack initiation and growth. Fatigue modelling should therefore be calibrated on the specific material/process state under a justified oral analogue.

Wear, interfaces, tribocorrosion and ageing

Dental wear is a contact-mechanics problem in which material loss and surface evolution change the future stress field. Two-body wear involves direct contact between opposing surfaces, whereas three-body wear includes particles/debris between them. Impact-sliding in chewing simulators can combine indentation, tangential traction and repeated plastic or brittle deformation. A simplified Archard relationship is often useful as a first engineering descriptor:

$$V = k\,W\,L\,/\,H \qquad (6)$$

where V is wear volume, W normal load, L sliding distance, H hardness and k an empirical wear coefficient. However, contemporary dental data repeatedly show why k cannot be treated as a material constant. Ghabchi et al. [27] compared zirconia, ceramic and printed nano-hybrid resin crowns; Li et al. [28] showed a distinctive filled-PEEK wear mechanism under simulated chewing; Osiewicz et al. [30] separated two- and three-body wear of BPA-free composites; and Wu et al. [29] examined wear resistance of repaired composites after simulated oral ageing. The tribosystem-antagonist, lubricant, debris, contact pressure, roughness and motion-must therefore be reported completely.

A particularly important clinical-mechanical interaction is antagonist wear. A hard restorative surface can be mechanically durable while damaging opposing enamel if roughness or fractured grains increase abrasivity. Coatings and polishing should consequently be evaluated with at least two responses: restoration wear/fatigue and antagonist loss. da Silva et al. [19] explicitly combined fatigue life and antagonist wear for coated ultra-translucent zirconia, an approach that is well suited to multi-objective engineering optimization.

Many restorations are layered systems rather than monolithic materials. A resin-cemented ceramic, resin-dentin interface or Ti-base/hybrid abutment transfers load across interfaces whose stiffness, thickness, roughness and chemistry affect the global stress state. Static bond strength can miss progressive hydrolytic

damage and mixed-mode crack growth. Stape et al. [44] specifically examined fatigue strength of resin-dentin interfaces after prolonged ageing, while Chen et al. [48] studied fatigue of direct and CAD/CAM restoratives bonded to a dentine-analogue substrate. These studies support interface fatigue as a distinct doctoral topic rather than a secondary bonding test.

From an engineering standpoint, interface durability can be framed using cohesive-zone parameters-peak traction, fracture energy and damage evolution-or experimentally through double-cantilever, notched, push-out or cyclic bonded-specimen configurations appropriate to the geometry. The key is to connect interface microstructure and ageing to an energy-based failure parameter that can be transferred into FEA. This would directly improve the predictive value of crown, veneer and implant-abutment simulations.

Titanium derives corrosion resistance from a passive oxide film, but repeated contact and micromotion can mechanically disrupt this film while the electrochemical environment controls repassivation. Tribocorrosion therefore creates a synergistic damage term. A useful accounting identity is:

$$V_{total} = V_{mechanical} + V_{corrosion} + V_{synergy} \quad (7)$$

The final term is precisely why independent wear and corrosion tests are insufficient. Vegas-Bustamante et al. [37] directly examined tribocorrosion effects on titanium implant mechanical behavior, while Hariprasad et al. [39] synthesizes bio-tribocorrosion in the oral environment. Mangal et al. [38] highlights metal-ion release when titanium/titanium-alloy implants are coupled with dissimilar abutment materials, and Boffano et al. [40] reviews corrosion features across titanium alloys. A Mechanical Engineering thesis can contribute by synchronizing tangential force, coefficient of friction, electrochemical current/potential, wear depth and fatigue damage in one test platform.

Geometry changes are equally important. Implantoplasty removes material and changes surface topography, creating a combined section-loss and fatigue problem. Padullés-Roig et al. [42] combined fatigue experiments with FEA after implantoplasty, providing an example of validated computational mechanics. Future work can extend this by including realistic defect distributions, thread geometry, preload loss, corrosion-fatigue and patient-specific load spectra.

Accelerated ageing is valuable only if its damage mechanism corresponds to service degradation. Thermal cycling can accelerate interfacial mismatch; water storage can plasticize polymers and hydrolyze interfaces; ethanol may exaggerate matrix softening; and simultaneous mechanical loading can produce interaction effects not captured by sequential protocols. Troha et al. [32] provides a particularly important caution by comparing real-time and accelerated ageing: 12-month real-time storage did not map trivially onto the accelerated protocols for all mechanical responses. This means that accelerated tests should be calibrated rather than assumed equivalent to a fixed number of clinical years.

For printed resins, ageing interacts with conversion and layer architecture. Alhotan et al. [49] reported material-dependent property changes after thermal cycling, while Alhotan et al. [24] studied high-impact printed denture resins after thermomechanical ageing. Ateş and Gungormus [35] showed that CAD/CAM polymer systems respond differently to thermal ageing in surface and mechanical outcomes. A robust PhD can therefore compare sequential versus simultaneous ageing, identify a damage marker such as modulus loss or crack density, and construct a time-temperature-moisture-fatigue superposition model where scientifically justified.

Manufacturing, post-processing and experimental mechanics

Manufacturing should be treated as part of the material definition. A ‘zirconia’ or ‘resin’ specimen without a complete process history is an incompletely specified engineering material. For CAD/CAM milling, tool condition, bur wear, machining direction, coolant, minimum wall thickness, finishing and polishing determine surface integrity. Pilecco et al. [46] demonstrates that milling protocol can influence fatigue behavior; this motivates direct measurement of machining damage rather than treating it as uncontrolled scatter.

For vat-polymerized resins, a minimum reporting set should include printer/technology, resin lot, build orientation, layer thickness, support strategy, exposure

settings if available, wash solvent/time, post-cure device, time/temperature, storage before test and ageing condition. Rizzante et al. [6], Mudhaffer et al. [7], Korkmaz et al. [9] and Flottes et al. [14] show that these variables are capable of shifting physical/mechanical outcomes. The engineering goal is to convert them into quantitative process maps that predict conversion, shrinkage, anisotropy, surface condition and fatigue life.

For zirconia AM, print settings interact with debinding and sintering. Orientation and layer thickness influence flaw orientation; sintering shrinkage and thermal gradients influence density, distortion and microstructure; support removal and surface finishing create additional damage. Recent systematic and experimental literature is sufficiently mature to justify full factorial or response-surface DOE on a limited set of high-impact variables, but the outputs should include fatigue/reliability-not only flexural strength [11, 12].

Surface engineering creates another multi-objective design space. Air abrasion may increase micromechanical retention but introduce flaws; polishing can reduce antagonist wear but alter bonding; coatings may improve bioactivity or antimicrobial behavior but introduce an adhesion interface; laser texture can change roughness, residual stress and wettability. Cantelli et al. [2] and da Silva et al. [19] demonstrate why a surface treatment should be evaluated under fatigue and wear. For implants, any surface intervention should additionally be assessed under corrosion/tribocorrosion.

Reproducible experimental mechanics

A coherent PhD experimental program should be designed around a causal hypothesis. If the hypothesis is that print orientation changes fatigue because it changes interlayer flaw orientation, then the minimum evidence chain is orientation -> quantified defect/anisotropy -> static response -> fatigue -> fracture origin. If the hypothesis is that a coating improves zirconia clinical durability, the chain is coating process -> roughness/adhesion/residual stress -> fatigue/wear -> antagonist response -> failure mode. Mechanical testing without microstructural or failure characterization rarely establishes causality.

Specimen design must balance standardization and clinical relevance. Rectangular bars/discs are useful for material screening because stress fields are interpretable; anatomically shaped crowns and implants are useful for system-level validation but introduce geometry and contact complexity. A strong doctoral design uses both: standardized specimens to identify material/process mechanisms, then component-level specimens to validate transferability. Replicate numbers should be justified by expected scatter and statistical model, especially for brittle/Weibull or fatigue survival analysis.

Fractography is not optional when the research question concerns fracture. SEM or high-quality optical fractography can identify crack origin, flaw type, hackle/river patterns, filler pull-out, interlayer failure or coating delamination. Surface profilometry or 3D optical metrology should quantify roughness and wear volume rather than relying only on mass loss. For printed materials, uCT can characterize internal pores non-destructively and create a direct bridge to defect-informed simulation. For composites, FTIR/Raman and DMA can connect conversion/network state to mechanical response. For metals, electrochemical monitoring synchronized with tribological force provides a mechanistic tribocorrosion signature.

The most useful experiment is one in which the boundary conditions are recorded well enough to reproduce the test in FEA. Contact geometry, friction, support compliance, load location, specimen dimensions and material constitutive data should therefore be treated as model inputs from the start. This reduces the common disconnect between a laboratory paper and a separate numerical paper that uses idealized assumptions.

A major reason dental-material studies are difficult to compare is incomplete reporting of variables that strongly influence mechanics. Reproducibility requires more than naming the test machine and specimen material. For fatigue, the report should include specimen geometry, surface preparation, support stiffness, indenter/contact geometry, waveform, frequency, load ratio, loading schedule, environment, failure definition, run-out and statistical treatment. For wear, it should include antagonist material and radius, normal load, motion path, frequency/speed, lubricant, temperature, total sliding distance, debris handling

and the method used to calculate wear. For additive manufacturing, printer/process type, orientation, layer thickness, support strategy, wash/post-cure or debind/sinter history, surface finishing and conditioning should accompany the mechanical result. The heterogeneity highlighted by Flottes et al. [14], Dascanio et al. [15], Pinelli et al. [10] and Ruggiero et al. [11] makes such reporting essential.

Uncertainty should also be treated as data. Dimensions, load-cell accuracy, roughness, temperature, material batch and alignment each contribute to experimental scatter. A high-quality doctoral study should report repeatability of the measurement system, justify sample size from the targeted effect or reliability analysis, identify outliers using predefined criteria rather than post-hoc deletion, and preserve specimen-level data for survival or regression analysis. Larrañaga-Ordaz et al. [50] makes reproducibility itself an explicit scientific topic for accelerated dentin-composite fatigue, demonstrating that test-system variability can be a meaningful research outcome rather than background noise.

Mechanism confirmation is the final reporting layer. A strength or fatigue number should be accompanied by evidence that the intended damage mechanism occurred. Fractography should identify crack origin and propagation region; wear testing should distinguish material loss from transferred debris or antagonist wear; tribocorrosion should track electrochemical response during mechanical contact; and interface fatigue should determine whether failure is adhesive, cohesive or mixed. Without mechanism confirmation, statistical differences may reflect unintended test artifacts.

Fractography and metrology-informed failure analysis

Failure analysis is the bridge between mechanical testing and mechanistic modelling. For ceramics, fracture-surface examination should be planned before the test so that pieces are recovered, orientation is preserved and the loaded/contact surface is traceable. The objective is to locate the origin, determine whether it coincides with a pore, machining groove, margin, interface or contact damage zone, and then relate the origin to the calculated tensile stress. This origin-specific workflow is more informative than classifying a specimen only as 'fractured'. Cantelli et al. [2], da Silva et al. [19], Packaeser et al. [3], Pilecco et al. [46] and the 2026 veneer studies collectively support the importance of surface, contact and geometric conditions in ceramic fatigue.

For additively manufactured materials, metrology should be three-dimensional whenever possible. A 2D polished cross-section can miss connected porosity, oriented interlayer defects or subsurface voids. uCT can quantify pore size, sphericity, orientation and location; surface profilometry or optical scanning can quantify roughness and waviness; SEM can characterize local morphology; and XRD can quantify crystalline phase changes in ceramic systems. These descriptors can then be transformed into model inputs or predictors. Bergler et al. [13], Sabatini et al. [12] and Wu et al. [45] illustrate the value of linking structural characterization to mechanical response rather than treating microstructure as a decorative image.

For composites and interfaces, the relevant morphology includes filler exposure, matrix plucking, interfacial gaps, fiber bridging/pull-out and cohesive versus adhesive fracture. For titanium systems, pits, fretting scars, oxide-film disruption and crack origins should be co-located with electrochemical history when possible. The result is a failure atlas that can support mechanism-based clustering of specimens and reduce the risk of fitting one life model to multiple unrelated failure modes.

Computational mechanics, reliability and life prediction

FEA is most valuable when it tests a mechanistic hypothesis and is validated against experiment. Dental models frequently contain strong simplifications: isotropic and homogeneous tissues, perfectly bonded interfaces, static point loads and elastic materials. These assumptions can be suitable for sensitivity studies but can also conceal the mechanisms that determine fatigue or interface failure. The 2025 dynamic framework study by Sahin Hazir et al. [16] and the hybrid-abutment study by Poovarodom et al. [17] demonstrate how material modulus redistributes stress; the next step is to validate such predictions with measured strain, displacement or component failure.

For ceramics, defect-sensitive FEA can be combined

with Weibull or weakest-link statistics. For thin veneers/crowns, local tensile stress at the cementation surface can be mapped against defect distributions inferred from fractography. For composites and interfaces, cohesive-zone models can represent progressive debonding, while viscoelastic constitutive models can capture time-dependent polymer response. For printed anisotropic materials, orthotropic elastic or fatigue properties should be measured by orientation instead of assuming isotropy. For implants, nonlinear contact and screw preload are essential if the question concerns micromotion, fretting or connection fatigue.

Life prediction should include uncertainty. Manufacturing scatter, flaw size, contact location and patient load variation can be represented through Monte Carlo simulation, Latin hypercube sampling or probabilistic design. Rather than report a single peak stress, a thesis can predict probability of failure versus cycles and validate the distribution experimentally. This moves dental FEA from explanatory color maps toward engineering reliability.

Machine learning can assist once a physically meaningful dataset exists. DOE/response-surface models are often preferable for small experimental datasets because they expose interactions and uncertainty. For larger multi-factor datasets, Gaussian-process regression, random forests or neural networks can serve as surrogate models for expensive FEA/experiments. The important principle is physics-informed feature selection: print orientation, pore fraction, modulus, roughness, contact stress and ageing dose are more interpretable predictors than arbitrary labels. Multi-objective optimization can then seek a Pareto set balancing fatigue life, wear, antagonist loss, stress shielding, manufacturability and cost.

Verification, validation and uncertainty quantification

FEA becomes predictive only when verification and validation are separated. Verification asks whether the numerical equations are being solved correctly: mesh refinement, element quality, contact stabilization, time increment and solver tolerances should not materially change the target output. Validation asks whether the model represents the physical system: predicted strain, displacement, contact area, failure location or load should be compared with independent experiments. A model may be numerically converged but physically wrong if the cement layer, boundary restraint, screw preload, support modulus or friction coefficient is unrealistic.

The most efficient validation experiment is not always a full anatomical fatigue test. A hierarchy can be used. Level 1 validates material constitutive data on coupons. Level 2 validates a simplified contact or bonded geometry with DIC/strain gauge/displacement. Level 3 validates the anatomical crown, veneer, abutment or implant component under the same load path used in the model. Level 4 validates durability by comparing failure location and life distribution. Sahin Hazir et al. [16], Poovarodom et al. [17] and Padullés-Roig et al. [42] provide a contemporary basis for moving implant modelling toward such multi-level validation.

Uncertainty quantification is particularly important for brittle ceramics and patient-specific geometry. A deterministic maximum stress does not express variation in flaw size, thickness, contact position, modulus, cement thickness or load. Monte Carlo or Latin-hypercube sampling can propagate these uncertainties, while Gaussian-process surrogates can reduce the computational cost of repeated nonlinear simulations. The final output should be a confidence or failure-probability band rather than a single stress value whenever scatter is intrinsic to the material or process.

Model calibration should avoid circularity. If the same failure data are used to fit and validate the model, agreement is expected and does not establish predictive capability. A doctoral project should reserve an independent material orientation, thickness, surface treatment, geometry or load case as a validation set. This is particularly valuable in additive manufacturing, where a process model may be calibrated at 0 and 90 degree orientations and challenged at 45 degrees, or calibrated on coupon fatigue and challenged on a crown component.

Evidence synthesis and cross-domain gaps

The selected 50-paper corpus is used as an engineering evidence map rather than a numerical quality ranking. The variable-mechanism-gap synthesis in Table 4 shows that the literature is sufficiently mature to formulate focused research questions while still leaving major

gaps in protocol standardization, long-term fatigue, coupled degradation and experimental validation.

The 2025 evidence establishes the principal process-to-damage variables. Buser et al. [1] showed that loading protocol can alter the apparent reliability of lithium disilicate and 5Y-TZP; Cantelli et al. [2] linked zirconia surface conditioning to bonded fatigue survival; Packaeser et al. [3] demonstrated sensitivity to core material and occlusal contact; and Pilecco et al. [46] showed that milling protocols can shift CAD/CAM fatigue response. For digital materials, Pinelli et al. [10] reported generally higher flexural strength and Weibull reliability for milled than printed zirconia, while Rizzante et al. [6], Mudhaffer et al. [7] and Alhotan et al. [49] demonstrated the roles of polymerization, orientation and thermal cycling in printed resins. Doumit et al. [8] supplied clinically relevant 24-month wear evidence. At the implant scale, Schettini et al. [36], Vegas-Bustamante et al. [37], Sahin Hazir et al. [16] and Poovarodom et al. [17] collectively connect manufacturing route, tribocorrosion and stiffness-dependent stress redistribution to component durability. These studies support a common interpretation: a reported strength or wear value is meaningful only when its process history, contact state and ageing history are defined.

The 2026 studies sharpen the same engineering direction and expose where validation remains immature. Bakitian [18], Zurlo Dellazzana et al. [4], Prott et al. [5] and Spitznagel et al. [43] show that multilayer architecture, thickness and restoration configuration interact with ceramic fatigue. Ruggiero et al. [11], Sabatini et al. [12], Bergler et al. [13] and Wu et al. [45] reinforce the need to report print orientation, layer/firing history and process-induced microstructure when evaluating additive zirconia. For definitive resins, Flottes et al. [14], Dascanio et al. [15], Terán Lozada et al. [25], Alhotan et al. [24], Korkmaz et al. [9] and Benalcázar-Jalkh et al. [26] collectively show material- and process-dependent property envelopes but limited long-term standardization. Watanabe et al. [33], Stape et al. [44], Troha et al. [32] and Osiewicz et al. [30] extend the problem toward reinforcement architecture, interface fatigue, real-time versus accelerated ageing and two-/three-body wear. Hariprasad et al. [39], Boffano et al. [40] and Padullés-Roig et al. [42] further emphasize coupled electrochemical-mechanical implant degradation and validated FEA. Thus, the strongest cross-domain evidence favors mechanism-specific, state-variable-based testing followed by component-level validation rather than direct comparison of isolated property values.

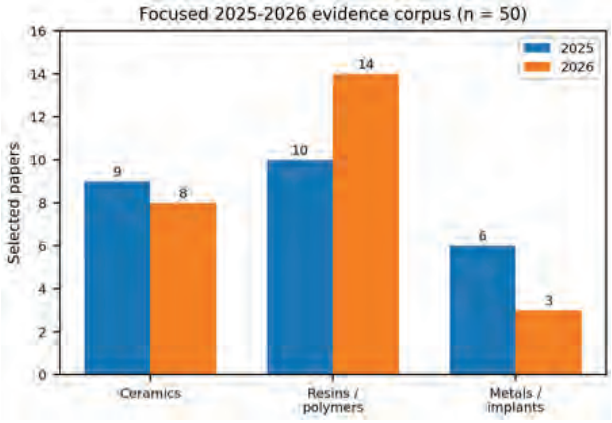


Fig. 4. Distribution of the focused 50-paper evidence corpus: 25 papers from 2025 and 25 from 2026

Table 4. Condensed 2025-2026 evidence synthesis and unresolved gaps

Domain	Representative evidence	Unresolved engineering need
Ceramics / zirconia	Protocol, contact, thickness, milling and surface condition change fatigue [1, 2, 3, 46, 5, 43, 4]	Standardized fatigue + defect-sensitive thickness/reliability models
AM zirconia	Printed-vs-milled differences; orientation, layer and firing effects [13, 10, 11, 12, 45]	Process-defect-life mapping using micro-CT/XRD/fractography + fatigue
Printed resins / composites	Orientation, cure and ageing strongly affect response [49, 24, 23, 26, 22, 15, 14, 9, 7, 6, 25]	Long-term fatigue/clinical validation and standardized print-to-test reporting
PEEK / polymer CAD/CAM	Wear and thermal ageing are material/tribosystem dependent [35, 28]	Coupled wear-fatigue and surface/interface design
Metals / implants	Manufacturing, ion release, corrosion and tribocorrosion interact [40, 39, 41, 38, 42, 36, 37]	Synchronized bio-tribocorrosion-fatigue + validated nonlinear contact FEA

The first cross-domain gap is protocol standardization. Fatigue studies differ in specimen geometry, support stiffness, waveform, frequency, load increment, thermocycling, failure definition and run-out. Wear studies differ in antagonist, lubrication, contact force, sliding path, debris and reported metric. Additive-manufacturing studies vary in orientation, post-cure or sintering and surface finishing. These differences are not merely reporting inconveniences; they change the damage mechanism. Dascanio et al. [15] and Flottes et al. [14] make this particularly clear for 3D-printed

resins, while Pinelli et al. [10] and Ruggiero et al. [11] expose the same challenge for printed zirconia.

The second gap is the weak bridge between accelerated laboratory ageing and actual long-term degradation. Thermocycling and water storage are widely used, yet Troha et al. [32] shows that real-time and accelerated ageing can produce different mechanical outcomes. Doumit et al. [8] is valuable because it provides clinical wear data that can calibrate laboratory models, but longer follow-up and matched laboratory conditions are needed. A major research opportunity is to define equivalent-damage metrics instead of converting arbitrary cycle counts into 'years'.

The third gap is coupled degradation. Tribocorrosion studies explicitly couple wear and electrochemistry, but many ceramic/polymer studies still separate fatigue, wear and ageing. In service, a roughened worn surface can raise contact stress; water can accelerate interface fatigue; crack formation can increase debris; and corrosion pits can become fatigue initiators. Future research should deliberately test interaction terms using factorial designs and simultaneous exposure where possible.

The fourth gap is model validation. Dental FEA has progressed beyond simple static comparisons, but predictive life modelling remains limited. Dynamic analyses such as Sahin Hazir et al. [16], hybrid-abutment stress studies such as Poovarodom et al. [17], and fatigue-FEA integration such as Padullés-Roig et al. [42] provide a foundation. The next research generation should report mesh convergence, contact assumptions, material data sources, uncertainty and direct comparison with measured strains/displacements/failure sites.

The fifth gap is translation from material coupon to dental component. Standard flexure or wear specimens are necessary for mechanism identification, but final validation must account for anatomical geometry, cement layer, substrate compliance, connector dimensions and patient-like contact. A two-stage approach-coupon mechanism discovery followed by component validation-is recommended.

RESEARCH ROADMAP AND INTEGRATED METHODOLOGY

A defensible doctoral topic should begin with an engineering failure mode, not with a material brand. The sequence recommended in Fig. 5 starts from an application-level problem (for example premature crown fatigue, printed-resin wear or implant tribocorrosion), selects a material/process system, formulates a mechanistic hypothesis, designs a controlled experiment, characterizes the resulting failure mechanism, builds a physics-based or statistically informed model, validates it and ends with an actionable design/process rule. This sequence prevents the thesis from becoming a catalogue of property tests.

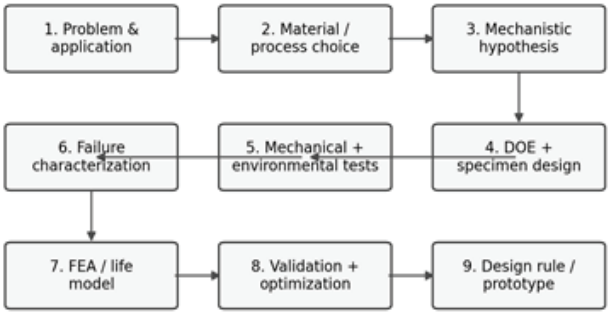


Fig. 5. Proposed translation path from a dental-material failure problem to validated life prediction and an actionable process/design rule

The research scope can be organized into six streams. Stream A-damage-tolerant ceramics-focuses on zirconia/lithium-disilicate thickness, gradient location, surface treatment, AM defects and probabilistic fatigue. Stream B-additively manufactured polymers-focuses on orientation, layer thickness, conversion/post-cure, water/thermal ageing, fatigue and wear. Stream C-composite microstructure-focuses on filler loading, fiber architecture, shrinkage stress, interface ageing and repair. Stream D-PEEK/high-performance polymers-focuses on wear, creep/fatigue, bonding and stress redistribution. Stream E-metallic implants-focuses on fatigue, fretting, corrosion, tribocorrosion and surface/alloy engineering. Stream F-computational durability-focuses on validated dynamic/contact FEA, probabilistic life prediction and multi-objective optimization.

A Mechanical Engineering researcher does not need every possible instrument in one laboratory. A practical program can place UTM/fatigue/tribology/FEA/DOE

in the engineering department, use a dental collaborator for clinically realistic geometry and specimen preparation, and use shared facilities for SEM, XRD, uCT, electrochemistry or spectroscopy. The novelty should come from the integration of measurements into a mechanistic model, not from owning the equipment.

Table 5. Research-gap-to-method roadmap for Mechanical Engineering doctoral work

Research gap	Minimum strong experiment	Model / deliverable
Service-representative fatigue	Constant/step-stress + ageing + fractography	S-N/Weibull/survival life curve
AM anisotropy / defects	Multi-orientation specimens + micro-CT/SEM	Orthotropic/defect-informed FEA + process-life map
Accelerated ageing calibration	Matched real-time and accelerated cohorts	Equivalent-damage kinetics with uncertainty
Wear-fatigue interaction	Sequential and simultaneous wear-fatigue	Contact + remaining-life model
Tribocorrosion-fatigue	Potentiostat-tribology plus cyclic load	Synergy coefficient / coupled damage law
Interface fatigue	Aged cyclic bonded specimens	Cohesive-zone fatigue parameters
Patient-specific validation	Anatomical component test + DIC/strain	Validated nonlinear dynamic FEA
Multi-objective design	DOE across material/process/surface variables	Surrogate/RSM + Pareto process/design window

Recommended thesis architecture and practical scope

A four-work-package thesis architecture can accommodate most topics. WP1 establishes baseline material/process characterization and identifies dominant defects. WP2 performs service-representative fatigue, wear, ageing or tribocorrosion experiments using DOE. WP3 develops the computational/reliability model and calibrates it using WP1-WP2 data. WP4 validates the model on an anatomical or component-level prototype and performs multi-objective optimization. The thesis contribution is then a validated process/design rule rather than a collection of independent experiments.

WP1 - Process and structure: establish material chemistry, processing window, surface state, defect statistics and baseline mechanical properties.

WP2 - Damage and durability: apply fatigue/wear/ageing/tribocorrosion under controlled oral analogues and identify crack/wear/corrosion mechanisms.

WP3 - Model and reliability: develop FEA, cohesive, contact, fatigue, Weibull or degradation model with uncertainty quantification.

WP4 - Validation and optimization: test crown/veneer/abutment/implant prototypes, quantify model error, and derive a Pareto-optimal design/process window.

Practical scope selection

If access is limited to UTM, tribometer, optical microscopy and ANSYS/Abaqus, a strong thesis can still be built around printed/milled polymer wear-fatigue or ceramic thickness/contact mechanics with outsourced SEM. If a fatigue machine and chewing simulator are available, crown/veneer/abutment life becomes feasible. If an electrochemical workstation is available, titanium tribocorrosion is a high-value extension. If uCT or XRD is available through collaboration, additive zirconia process-defect research becomes especially strong. The equipment should follow the hypothesis; it should not determine the hypothesis.

For collaboration, the dental partner should define clinically meaningful geometry, preparation, cementation, antagonist and failure criteria, while the Mechanical Engineering team should own the load spectrum, metrology, mechanics, statistics and modelling. This division preserves clinical relevance without shifting the doctorate away from engineering. A materials/metallurgy partner can support SEM/XRD/uCT and an electrochemistry laboratory can support corrosion/tribocorrosion. The resulting interdisciplinary structure supports clinically relevant specimen design while retaining the thesis emphasis on Mechanical Engineering.

Research-topic selection and translational closure

Topic selection should balance scientific gap, equipment access, specimen complexity and the ability to produce a mechanistic contribution. A researcher with a UTM and access to a dental 3D printer can already investigate orientation/post-cure/ageing effects, but the PhD becomes substantially stronger if the study measures fatigue or residual strength and validates anisotropic FEA. Access to a tribometer opens wear-fatigue projects in PEEK, composites and printed resins; access to an electrochemical workstation enables titanium tribocorrosion; access to uCT enables defect-sensitive additive-zirconia reliability; and access to DIC strengthens validation of patient-specific models. The decision should therefore be made from

the required mechanism backward to the equipment, not from available equipment forward to a convenient but weak question.

The highest-value doctoral questions usually contain one controllable process variable, one service/environment variable and one dominant damage mechanism. Examples are print orientation x ageing -> fatigue of permanent crown resin; milling protocol x surface finish -> ceramic flaw population and fatigue; pH/fluoride x micromotion -> titanium tribocorrosion-fatigue; filler content x ageing -> composite fatigue/wear; or implantoplasty geometry x load spectrum -> implant fatigue. Adding too many materials and too many variables can dilute mechanism resolution. DOE should be used to select a small number of high-information combinations rather than building an unmanageably large material catalogue.

A final criterion is translational closure. The project should end with something an engineer or clinician can use: a minimum thickness curve, allowable process window, fatigue reliability map, post-curing recommendation, surface-treatment Pareto front, equivalent-ageing relation, tribocorrosion damage law, or validated simulation workflow. This deliverable distinguishes doctoral engineering research from comparative material screening.

Ready-to-develop project blueprints

Blueprint A - Additively manufactured permanent crown resin: Hypothesis: print orientation, layer architecture and post-curing create an anisotropic network whose ageing response controls fatigue more strongly than initial flexural strength. Use a compact DOE with orientation and cure as manufacturing factors and water/thermal ageing as the environmental factor. Measure conversion-sensitive properties, modulus, flexure, step-stress or constant-amplitude fatigue, roughness and fracture origin. Fit an orientation-dependent fatigue model and validate it on an anatomical crown. The literature of Mudhaffer et al. [7], Rizzante et al. [6], Alhotan et al. [49], Korkmaz et al. [9], Flottes et al. [14], Dascanio et al. [15] and Benalcázar-Jalkh et al. [26] provides a strong contemporary evidence base.

Blueprint B - Defect-tolerant additive zirconia: Hypothesis: the performance gap between printed and milled zirconia is governed by the orientation, size and location of process-induced flaws rather than by manufacturing route alone. Vary orientation/layer and one firing or sintering parameter; quantify density/porosity and phase/microstructure; measure flexural strength, Weibull reliability and fatigue; locate fracture origins; and develop a defect-informed FEA/probabilistic model. Validate on a thin crown or veneer. Pinelli et al. [10], Ruggiero et al. [11], Sabatini et al. [12], Wu et al. [45] and Bergler et al. [13] directly support this roadmap.

Blueprint C - Minimum-thickness design of minimally invasive ceramics: Hypothesis: a universal minimum thickness is mechanically invalid because contact, support and material-specific crack resistance interact. Compare a tightly selected ceramic pair across clinically meaningful thicknesses and contact configurations; perform thermomechanical fatigue and fractography; calculate intaglio/contact tensile stress; and build a probability-of-failure map. Zurlo Dellazzana et al. [4], Prott et al. [5], Spitznagel et al. [43], Buser et al. [1] and Packaeser et al. [3] provide the foundation.

Blueprint D - Wear-fatigue coupling in PEEK and polymeric CAD/CAM materials: Hypothesis: wear-induced topography and section loss alter subsequent fatigue life, and the effect depends on polymer/filler architecture. Run controlled two-body or chewing-simulator wear under artificial saliva, quantify 3D wear and antagonist loss, then measure residual fatigue life. Model evolving contact geometry and compare predicted stress change with residual life. Li et al. [28], Ateş and Gungormus [35], Osiewicz et al. [30] and Wu et al. [29] motivate this coupled rather than isolated tribology study.

Blueprint E - Resin-dentin interface durability: Hypothesis: long-term adhesive failure is controlled by ageing-induced degradation of interface fracture resistance, not merely initial bond strength. Select a small number of adhesive strategies, conduct real-time/accelerated ageing, cyclic interface tests and failure-mode analysis, and identify cohesive-zone traction-separation/fatigue parameters. Validate the model on a bonded restoration geometry. Stape et al. [44], Larrañaga-Ordaz et al. [50], Chen et al. [48] and Tsujimoto et al. [31] frame the need.

Blueprint F - Titanium bio-tribocorrosion-fatigue: Hypothesis: mechanically assisted passive-film disruption produces a synergistic surface-damage state that reduces residual fatigue life beyond independent corrosion or wear. Use artificial saliva with controlled pH/fluoride, synchronized reciprocating micromotion and electrochemical measurement, quantify synergy, then fatigue pre-damaged specimens or implant components. Correlate pits/scars with crack origins. Vegas-Bustamante et al. [37], Mangal et al. [38], Herbster et al. [41], Hariprasad et al. [39] and Boffano et al. [40] establish the contemporary problem.

Blueprint G - Surface-engineered zirconia with multi-objective durability: Hypothesis: surface treatments that improve bonding or functional coating behavior can simultaneously alter flaw severity and antagonist wear. Compare one baseline and a limited number of surface/coating routes; measure roughness, surface damage, bond-fatigue, crown fatigue and antagonist wear; use Pareto optimization rather than selecting the treatment with the highest single property. Cantelli et al. [2] and da Silva et al. [19] provide direct fatigue/wear motivation.

Blueprint H - Experimentally validated patient-specific implant mechanics: Hypothesis: material stiffness and component geometry interact with preload/contact to redistribute stress and micromotion, and validated dynamic models can identify lower-risk configurations. Build CT/CAD-based geometry, characterize component materials, include contact and preload, perform DIC/strain-gauge or displacement validation, then add fatigue/reliability and uncertainty. Sahin Hazir et al. [16], Poovarodom et al. [17] and Padullés-Roig et al. [42] provide a natural progression from explanatory FEA toward predictive mechanics.

Four-year execution and publication logic

Year 1 should close the literature-to-method gap: finalize the research question, select one primary material system, conduct pilot characterization, establish specimen manufacturing repeatability and complete measurement-system validation. The first publication can be the present critical review or a more focused systematic review. The decision gate at the end of Year 1 is whether the selected process variables produce a measurable and mechanistically interpretable change rather than merely statistical noise.

Year 2 should generate the core experimental mechanics dataset. A DOE should vary only the factors justified by pilot data and the review. Fatigue, wear, ageing or tribocorrosion should be accompanied by surface/microstructural diagnostics and appropriate reliability statistics. The second publication should report a process-structure-damage relationship, not just property ranking. A third publication can focus on coupled degradation or mechanism-specific failure analysis if sufficient depth exists.

Year 3 should translate the experimental data into a computational or reliability model. Material inputs should be measured where possible; contact and boundary conditions should be verified; uncertainty should be propagated; and an independent validation condition should be reserved. The publication target is a validated mechanics/life-prediction paper. At this stage, the project should be able to predict an outcome that was not used during model fitting.

Year 4 should perform component-level validation and optimization. The model can be used to reduce the experimental design space, identify promising material-process-geometry combinations and define a process/design window. Final experiments should challenge the predicted optimum and at least one non-optimal control. The resulting contribution can be expressed as an engineering decision chart, reliability curve, Pareto front, validated digital workflow or component design guideline. This structure gives the thesis a coherent narrative from evidence to mechanism to prediction to design.

CONCLUSIONS

The current 2025-2026 literature shows that dental-material durability is controlled less by a single nominal property than by interactions among manufacturing, microstructure, geometry, surface condition, cyclic load and oral environment. Zirconia and glass-ceramics remain strongly flaw- and thickness-sensitive; additively manufactured zirconia introduces orientation, layer and firing variables that require reliability-based characterization. Printed resin systems are advancing rapidly but remain sensitive to orientation, post-curing and ageing, with long-term fatigue and clinical evidence still less mature than for established milled materials. Direct composites continue to reveal the importance of

matrix-filler-interface design, polymerization stress and ageing. PEEK and related polymers offer distinctive stress-transfer and wear behavior, while titanium/Co-Cr systems shift the durability problem toward fatigue, fretting, corrosion and tribocorrosion.

Five priorities should define the next Mechanical Engineering research phase: (1) service-representative and reproducible fatigue/wear protocols; (2) defect-informed process-structure-life models for additive manufacturing; (3) simultaneous or calibrated ageing rather than arbitrary cycle-to-year conversions; (4) experimentally validated dynamic/contact and probabilistic FEA; and (5) multi-objective optimization that balances strength, fatigue, wear, antagonist response, corrosion, manufacturability and clinical geometry. The most publishable and defensible PhD work will close the loop from controllable process variables to failure mechanism and validated design guidance.

The research roadmap proposed here provides that loop. A researcher can select a material family, identify a process or environmental variable, choose a dominant failure mechanism, design a mechanistic experiment, quantify reliability, build and validate a computational model, and finish with a process/design window. In this way dental materials become not an adjacent biomedical topic but a rigorous Mechanical Engineering research domain integrating materials science, solid mechanics, fracture/fatigue, tribology, manufacturing, surface engineering and computational design.

ACKNOWLEDGEMENT

The authors acknowledge the academic and institutional support associated with their affiliations.

REFERENCES

1. Buser, R., et al. (2025). Fatigue behavior and reliability of pressed lithium disilicate ceramics compared to 5Y-TZP zirconia under different loading protocols. Dental Materials, 41(2), 134-140. https://doi.org/10.1016/j.dental.2024.11.005
2. Cantelli, V., et al. (2025). Fatigue survival rate of resin-bonded zirconia-based ceramics after different surface treatments. Dental Materials, 41(9), 1113-1119. https://doi.org/10.1016/j.dental.2025.06.020
3. Packaeser, M. G., et al. (2025). Influence of core material and occlusal contact pattern on fatigue behavior of different monolithic ceramic crowns. Journal of the Mechanical Behavior of Biomedical Materials, 163, 106891. https://doi.org/10.1016/j.jmbbm.2025.106891
4. Zurlo Dellazzana, F., et al. (2026). Fatigue behavior of dental ceramics for occlusal veneers: Impact of the material thickness. Journal of the Mechanical Behavior of Biomedical Materials, 179, 107439. https://doi.org/10.1016/j.jmbbm.2026.107439
5. Prott, L. S., et al. (2026). Fatigue performance and failure load of minimally-invasive occlusal veneers made of lithium disilicate and composition-gradient multilayered zirconia: An in vitro study. Journal of Prosthodontic Research, 70(2), 233-242. https://doi.org/10.2186/jpr.JPR_D_25_00068
6. Rizzante, F. A. P., et al. (2025). Influence of polymerization protocols on the physical and mechanical properties of a 3D printed resin. The Journal of Prosthetic Dentistry, 133(4), 1091.e1-1091.e6. https://doi.org/10.1016/j.prosdent.2025.01.006
7. Mudhaffer, S., et al. (2025). Effects of print orientation and artificial aging on the flexural strength and flexural modulus of 3D printed restorative resin materials. The Journal of Prosthetic Dentistry, 133(5), 1345-1357. https://doi.org/10.1016/j.prosdent.2024.08.008
8. Doumit, M., et al. (2025). Wear behavior of 3D printed, minimally invasive restorations: Clinical data after 24 months in function. The Journal of Prosthetic Dentistry, 134(3), 662-669. https://doi.org/10.1016/j.prosdent.2025.03.014
9. Korkmaz, Y. N., et al. (2026). Effects of different post curing times and methods on fracture resistance and flexural strength of 3D-printed permanent crown resin. International Journal of Prosthodontics, online ahead of print, 1-21. https://doi.org/10.11607/ijp.9575
10. Pinelli, L. A. P., Ferreira, I., & Dos Reis, A. C. (2025). Analysis of flexural strength and Weibull modulus of printed and milled zirconia: A systematic review. The Journal of Prosthetic Dentistry, 134(3), 628.e1-628.e8. https://doi.org/10.1016/j.prosdent.2025.05.032
11. Ruggiero, M. M., et al. (2026). Is additive manufacturing of dental zirconia comparable to subtractive methods when considering printing orientation and layer thickness? A systematic review and meta-analysis. Journal of Esthetic and Restorative Dentistry, 38(3), 604-634. https://doi.org/10.1111/jerd.13514
12. Sabatini, G. P., et al. (2026). Flexural and structural behavior of additively manufactured zirconia: Influence of firing process and build orientation. The Journal of

Prosthetic Dentistry, 135(3), 612.e1-612.e8. https://doi.org/10.1016/j.prosdent.2025.10.043

13. Bergler, M., et al. (2026). Comparison of the flexural strength and microstructure of 3D printed and milled mono- and multilayer zirconia. Journal of the Mechanical Behavior of Biomedical Materials, 177, 107370. https://doi.org/10.1016/j.jmbbm.2026.107370

14. Flottes, Y., et al. (2026). Properties of 3D printed resins for definitive dental restorations: A systematic review. The Journal of Prosthetic Dentistry, 135(4), e6-e27. https://doi.org/10.1016/j.prosdent.2025.11.040

15. Dascanio, R., Kury, M., Özcan, M., & Cavalli, V. (2026). How durable are 3D-printed dental resin-based composites? An umbrella review. Journal of Dentistry, 168, 106626. https://doi.org/10.1016/j.jdent.2026.106626

16. Sahin Hazir, D., Sozen Yanik, I., Guncu, M. B., & Canay, R. S. (2025). Biomechanical behavior of titanium, cobalt-chromium, zirconia, and PEEK frameworks in implant-supported prostheses: a dynamic finite element analysis. BMC Oral Health, 25(1), 97. https://doi.org/10.1186/s12903-025-05486-5

17. Poovarodom, P., et al. (2025). Mechanical behavior of hybrid custom implant abutments with various crown materials: a 3D finite element analysis. BMC Oral Health, 25(1), 1106. https://doi.org/10.1186/s12903-025-06445-w

18. Bakitian, F. (2026). Influence of manufacturer, laboratory, clinical, and aging-related factors on the mechanical properties of multilayered zirconia: a systematic review of laboratory studies. BMC Oral Health, 26(1), 988. https://doi.org/10.1186/s12903-026-08176-y

19. da Silva, A. C., et al. (2025). Ultra-translucent zirconia crowns with antimicrobial glass coatings: fatigue life and antagonist wear analysis. Dental Materials, 41(11), 1376-1387. https://doi.org/10.1016/j.dental.2025.07.016

20. Ribeiro, M. T. H., et al. (2025). Physicomechanical properties of resin-based composites photo-polymerized using laser, polywave and quadwave curing lights. Dental Materials, 41(6), 699-707. https://doi.org/10.1016/j.dental.2025.04.001

21. Mohammadi Ghayeghchi, M., Atai, M., & Nodehi, A. (2025). Addition-fragmentation chain transfer monomer in dental resins: Synthesis, characterization, and properties. Dental Materials, 41(11), 1442-1453. https://doi.org/10.1016/j.dental.2025.08.007

22. Coldea, A., et al. (2025). Mechanical properties of 3D printed denture base polymers. The Journal of Prosthetic Dentistry, 133(5), 1361.e1-1361.e8. https://doi.org/10.1016/j.prosdent.2025.02.009

23. Artopoulou, I. I., et al. (2025). Comparative laboratory study of 3D printed denture base resin materials. The Journal of Prosthetic Dentistry, 134(5), 1964.e1-1964.e9. https://doi.org/10.1016/j.prosdent.2025.08.007

24. Alhotan, A., et al. (2026). Mechanical performance of high-impact 3D-printed denture base resins subjected to thermomechanical aging. Dental Materials, online ahead of print. https://doi.org/10.1016/j.dental.2026.07.016

25. Terán Lozada, M. I., et al. (2026). Physicomechanical properties of 3D printed resins for producing definitive dental crowns. The Journal of Prosthetic Dentistry, 135(5), 963-971. https://doi.org/10.1016/j.prosdent.2025.07.017

26. Benalcázar-Jalkh, E. B., et al. (2026). Evaluation of the fatigue behavior of implant-supported 3D-printed and milled resins for definitive crowns. Dental Materials, 42(5), 744-754. https://doi.org/10.1016/j.dental.2025.12.010

27. Ghabchi, B., et al. (2025). Wear behavior of CAD-CAM zirconia, ceramic, and 3D printed nano-hybrid resin crowns for the restoration of primary and permanent molars: An in vitro study. The Journal of Prosthetic Dentistry, 134(1), 177.e1-177.e8. https://doi.org/10.1016/j.prosdent.2025.02.040

28. Li, K., et al. (2025). Comparative study of wear properties of CAD/CAM PEEK materials, resin ceramics and dental enamel under simulated chewing conditions. Journal of the Mechanical Behavior of Biomedical Materials, 168, 107012. https://doi.org/10.1016/j.jmbbm.2025.107012

29. Wu, K., et al. (2026). Wear resistance of composite repairs: Direct vs. semi-direct techniques in simulated oral aging. Dental Materials, 42(9). https://doi.org/10.1016/j.dental.2026.04.010

30. Osiewicz, M. A., et al. (2026). Two- and three-body wear of BPA-free resin composites. Dental Materials, 42(7), 1172-1181. https://doi.org/10.1016/j.dental.2026.02.026

31. Tsujimoto, A., Thomas, J. C., & Caplan, D. (2025). Factors influencing the longevity of resin-based composite restorations. Dental Materials, 41(11), 1395-1401. https://doi.org/10.1016/j.dental.2025.07.018

32. Troha, L., et al. (2026). Filler amount influences long-term mechanical stability of experimental dental composites. Dental Materials, 42(6), 1020-1030. https://doi.org/10.1016/j.dental.2026.02.012

33. Watanabe, S., et al. (2026). Fatigue fracture resistance of a short fiber-reinforced flowable resin composite in single-layer and incremental-layer blocks. Dental Materials, 42(9), 1610-1618. https://doi.org/10.1016/j.dental.2026.04.011

34. Sari, O., et al. (2025). Mechanical behavior of subtractively and additively manufactured zirconia

and polyetheretherketone implant abutments after thermomechanical aging. The Journal of Prosthetic Dentistry, 134(3), 762.e1-762.e9. https://doi.org/10.1016/j.prosdent.2025.05.005

35. Ateş, G., & Gungormus, M. (2026). Effects of thermal aging on optical, surface, mechanical, and biological properties of CAD/CAM polymer-based dental materials. Polymers, 18(6), 760. https://doi.org/10.3390/polym18060760

36. Schettini, A., et al. (2025). Assessment of mechanical properties and microstructure of Co-Cr dental alloys manufactured by casting, milling, and 3D printing. The Journal of Prosthetic Dentistry, 133(3), 906.e1-906.e7. https://doi.org/10.1016/j.prosdent.2024.12.017

37. Vegas-Bustamante, E., et al. (2025). Effect of tribocorrosion on mechanical behavior of titanium dental implants: An in vitro study. Materials, 18(5), 1136. https://doi.org/10.3390/ma18051136

38. Mangal, M., et al. (2025). Metal ion release of titanium and titanium alloy dental implants with dissimilar abutment materials: A systematic review of in vitro studies. Journal of Indian Society of Periodontology, 29(3), 230-240. https://doi.org/10.4103/jisp.jisp_76_25

39. Hariprasad, A., et al. (2026). Bio-tribocorrosion of titanium dental implants in the oral environment: A narrative review. Cureus, 18(2), e103188. https://doi.org/10.7759/cureus.103188

40. Boffano, P., et al. (2026). Corrosion features of titanium alloys in dental implants: A systematic review. Journal of Maxillofacial and Oral Surgery, 25(1), 189-197. https://doi.org/10.1007/s12663-024-02169-6

41. Herbster, M., et al. (2025). Effects of Cr addition on Ti implant alloys (Ti-Cr/Ti-Al-V-Cr) to enhance corrosion and wear resistance. Journal of the Mechanical Behavior of Biomedical Materials, 163, 106899. https://doi.org/10.1016/j.jmbbm.2025.106899

42. Padullés-Roig, E., et al. (2026). Mechanical fatigue of titanium dental implants after implantoplasty: An in vitro study combined with finite element simulations. Journal of Functional Biomaterials, 17(5), 221. https://doi.org/10.3390/jfb17050221

43. Spitznagel, F. A., et al. (2026). Fatigue performance and failure loads of composition-gradient multilayer zirconia: Partial-coverage restorations versus single crowns at different thicknesses. Journal of Esthetic and Restorative Dentistry, 38(3), 539-552. https://doi.org/10.1111/jerd.70066

44. Stape, T. H. S., et al. (2026). On the degradation mechanics and fatigue strength of resin-dentin interfaces: Identifying weak links of adhesive strategies after prolonged ageing. Dental Materials, online ahead of print. https://doi.org/10.1016/j.dental.2026.04.016

45. Wu, Z., et al. (2026). Assessing mechanical properties and biosafety of dental zirconia for 3D printing and CNC milling. The Journal of Prosthetic Dentistry, 136(1), e369-e378. https://doi.org/10.1016/j.prosdent.2026.02.039

46. Pilecco, R. O., et al. (2025). The influence of milling protocols on the fatigue behavior of different CAD-CAM materials. Dental Materials, 41(9), 1179-1187. https://doi.org/10.1016/j.dental.2025.07.001

47. Karevan, Y., et al. (2025). Fatigue behavior of CAD-CAM composites versus lithium disilicate glass-ceramic. Dental Materials, 41(10), 1222-1230. https://doi.org/10.1016/j.dental.2025.07.004

48. Chen, Y., et al. (2026). Fatigue behaviour of direct and CAD/CAM indirect restorative materials bonded to a dentine analogue substrate. Dental Materials, online ahead of print. https://doi.org/10.1016/j.dental.2026.06.013

49. Alhotan, A., et al. (2025). Physical and mechanical properties of various resins for additively manufactured definitive fixed dental restorations: Effect of material type and thermal cycling. The Journal of Prosthetic Dentistry, 134(1), 238.e1-238.e8. https://doi.org/10.1016/j.prosdent.2025.02.042

50. Larrañaga-Ordaz, D., et al. (2026). Reproducibility of accelerated fatigue test of dentin-composite discs under cyclic diametral compression. Dental Materials, online ahead of print. https://doi.org/10.1016/j.dental.2026.03.157

DOI - 10.68648/ijte-q-2026-49-03-020

Manufacturing-Structure-Performance Engineering of Advanced Dental Materials: Additive Manufacturing, Surface Engineering and Computational Design

Jitendra B. Patil, Amol B. Patil
PhD Research Scholar
Mechanical Engineering
Srinivas University
Mangalore, Karnataka
✉ jitendrapatil11691@gmail.com
✉ amolbpatil11@gmail.com

Yogesh S. Andhale
Technical Lead
PTC Software India
Pune, Maharashtra
✉ yogesh.andhale87@gmail.com

ABSTRACT

Advanced dental materials are manufactured states whose performance depends on digital geometry, additive or subtractive processing, post-processing, surface modification, defects and oral loading. This focused critical review synthesizes 50 peer-reviewed publications from 2025–2026 using an explicitly documented search, eligibility, screening and evidence-coding protocol rather than a PRISMA systematic-review claim. The literature is integrated through a manufacturing–structure–performance (MSP) framework linking process variables to defect/microstructural and surface states and then to dimensional accuracy, fatigue, wear, adhesion and functional durability. Comparative synthesis shows that additive-manufacturing studies are strongest in static mechanical and accuracy characterization but weaker in defect-informed fatigue; surface-engineering evidence is increasingly multifunctional but coating durability is inconsistently tested; finite-element analysis is advancing toward patient-specific models but remains limited by idealized properties, contacts and experimental validation; and AI/ML is constrained by small, heterogeneous and weakly transferable datasets. A closed digital thread is proposed in which metrology and experiments calibrate predictive mechanics, validated simulations and tests generate traceable datasets, surrogate models accelerate prediction and multi-objective optimization identifies robust process/design windows. Priority directions include defect-informed zirconia AM, cure–conversion–fatigue modelling of printed resins, multifunctional coating durability, experimentally validated patient-specific FEA, inverse implant design, digital twins and sustainability-aware manufacturing optimization.

KEYWORDS *: Advanced dental materials; additive manufacturing; surface engineering; finite element analysis; intelligent optimization.*

NOTATIONS

AI – artificial intelligence; AM – additive manufacturing; CAD/CAM – computer-aided design/computer-aided manufacturing; DIC – digital image correlation; DLP – digital light processing; DOE – design of experiments; FDM/FFF – fused deposition modelling/fused filament fabrication; FEA – finite-element analysis; FTIR – Fourier-transform infrared spectroscopy; ML – machine learning; MOO – multi-objective optimization; MSP – manufacturing–structure–performance; micro-CT – micro-computed tomography; PBF – powder-bed fusion; PEEK – polyetheretherketone; RSM – response-surface methodology; SLA – stereolithography; X – process/material/contact-state vector; E – elastic modulus; K_IC – fracture toughness; H – hardness; P_f – probability of failure; σ_0 – Weibull characteristic strength; m – Weibull modulus; S_L – linear shrinkage; β – regression coefficient; ε – model residual/error term.

INTRODUCTION

Digital dentistry has changed the engineering definition of a dental material. A crown, denture

base, abutment or implant is no longer specified only by chemical composition; it is the output of a digital manufacturing chain that begins with image acquisition and CAD geometry and continues through toolpaths or sliced layers, machine settings, support strategy, thermal or photochemical post-processing, surface finishing and quality control. Each stage can change the final structure and surface state. Consequently, nominally identical material labels can produce different mechanical behavior when the manufacturing histories differ. Contemporary literature therefore needs an engineering framework that treats manufacturing history as part of the material state rather than as an incidental laboratory detail [1, 2].

This distinction is particularly important for additive manufacturing (AM). Layerwise fabrication creates geometric freedom and mass customization but also introduces anisotropy, interlayer interfaces, porosity, stair-step topography, support-related damage, conversion gradients or thermal residual stress depending on the process. For photopolymerized dental resins, orientation and post-curing directly influence final network formation and flexural behavior [3, 4]. For ceramic AM, printing direction, green-part architecture, debinding and sintering determine shrinkage, pore distribution and flaw orientation; recent meta-analyses continue to find that the performance of printed zirconia depends strongly on technology, orientation and layer thickness [5, 6]. Metallic dental components manufactured by powder-bed fusion add a different defect space involving energy density, lack-of-fusion pores, residual stress and rough surfaces that may later be altered by machining, polishing or coatings [7].

Surface engineering then becomes a second manufacturing layer. Polishing, air abrasion, laser texturing, anodization, acid treatment, hydrothermal processing, plasma spraying and multifunctional coatings are applied to change roughness, chemistry, wettability, bioactivity or antibacterial function. From a Mechanical Engineering standpoint, the surface cannot be optimized only for biological response because the same treatment can alter crack population, coating adhesion, friction, wear, corrosion and fatigue. Recent reviews of titanium and zirconia implant surfaces emphasize the increasing complexity of hierarchical micro/nano topographies and smart coatings, while experimental crown studies demonstrate the direct trade-off between functional coating, fatigue life and antagonist wear [8, 9, 10, 11].

Computational mechanics has developed in parallel. Finite-element analysis (FEA) is now widely used to explore implant geometry, framework stiffness, abutment design, contact and load transfer. Yet the value of a model depends on geometry fidelity, constitutive assumptions, contact definition, preload, loading spectrum, mesh convergence and validation. A 2026 systematic review of 278 implant FEA studies concluded that realism-enhancing practices such as patient-specific anatomy, anisotropic bone, nonlinear behavior, dynamic loading, fatigue, remodeling and uncertainty quantification remain incompletely adopted [12]. This means that a visually sophisticated FEA model can still be scientifically weak if its assumptions are unmeasured or its outputs are not validated against experiment.

Artificial intelligence (AI) creates a route to overcome some of these limits, especially when the objective is optimization rather than explanation. Machine-learning surrogate models can learn the input-output relation of thousands of simulations and provide predictions in seconds; finite-element-based machine learning has already been used to optimize complete-arch implant-supported prosthetic configurations [13]. In 2026, surrogate-assisted inverse design and time-resolved FEA-deep-learning frameworks further demonstrated the possibility of near-real-time patient-specific implant analysis [14, 15]. However, speed does not remove the need for validation. A fast surrogate reproduces the bias of the simulations and experiments used to train it; data quality, domain coverage and uncertainty therefore become engineering variables.

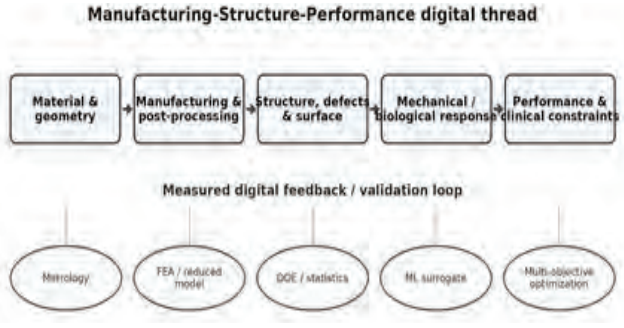


Fig. 1. Integrated manufacturing-structure-performance

digital thread linking measured state variables with FEA, data-driven prediction and optimization

This review is organized around the manufacturing–structure–performance (MSP) relationship. A controllable manufacturing or surface parameter changes the produced material state; that state alters mechanical, dimensional, tribological or biological response; and the response determines component performance. The central gap is that these links are usually studied in separate literatures—additive manufacturing, surface modification, computational mechanics and AI/ML—so a process improvement is rarely traced through measured structure to validated component-level performance. The review therefore contributes: (i) a transparent focused-review protocol for the 50-paper corpus; (ii) a common MSP causal framework for critical comparison; (iii) explicit appraisal of dataset heterogeneity, fatigue evidence, coating durability, model validation and AI/ML generalizability; and (iv) a closed-loop engineering architecture connecting metrology, FEA, surrogate prediction and robust multi-objective optimization. Fig. 1 summarizes the framework used throughout the manuscript.

REVIEW DESIGN, STUDY SELECTION AND EVIDENCE SYNTHESIS

Review scope, information sources and search strategy

The review was designed as a focused critical review, not a PRISMA systematic review or meta-analysis. To make evidence assembly auditable, the revised protocol specifies information sources, search concepts, eligibility rules and synthesis logic. Searches were structured around Scopus, Web of Science Core Collection, PubMed/MEDLINE, ScienceDirect and Google Scholar because these sources jointly cover dentistry, biomaterials, mechanical engineering, additive manufacturing and computational design. Boolean concepts combined (dental OR dentistry OR crown OR denture OR implant OR prosth*) with (additive manufactur* OR 3D print* OR CAD/CAM OR milling OR zirconia OR resin OR PEEK OR titanium OR cobalt-chromium) and, when relevant, (surface engineer* OR coating OR laser OR fatigue OR wear OR finite element OR FEA OR machine learning OR artificial intelligence OR optimization). Searches were restricted to peer-reviewed publications from 2025–2026 and supplemented by backward/forward citation checking.

Eligibility, screening and rationale for the 50-paper corpus

Studies were eligible when they addressed dental materials or dental components and reported, synthesized or modelled at least one explicit engineering link among manufacturing route/parameter, defect or microstructural state, surface condition, mechanical or dimensional response, FEA, AI/ML or optimization. Experimental studies, systematic reviews, meta-analyses, narrative reviews and computational studies were admissible. Exclusion criteria were non-dental applications without a transferable dental-material context, purely clinical reports without engineering variables, editorials/abstracts or non-peer-reviewed items, duplicate datasets without additional evidence, and papers with insufficient methods or inaccessible full text. Screening proceeded from deduplication to title/abstract relevance and then full-text eligibility.

The final corpus contains exactly 50 publications—25 from 2025 and 25 from 2026. This was a purposive quota-balanced design intended to provide a near-current two-year snapshot while preventing one publication year from dominating the comparison. The number 50 is therefore a depth–breadth design choice, not a claim of exhaustive capture: it enabled full-text coding of every included publication across manufacturing, structure/surface state, performance, validation and computation. Older foundational concepts are used as background but are not counted. Because the original focused search was not prospectively registered as PRISMA, intermediate attrition counts were not retained and are not reconstructed retrospectively.

Data extraction, evidence appraisal and synthesis

Each publication was coded using a common MSP template: material/component; manufacturing or surface route; controllable variables; measured defect/microstructure/surface descriptors; performance endpoints; aging/fatigue conditions; computational method; validation strategy; uncertainty/statistical treatment; and stated limitations. Evidence quality was interpreted through five engineering adequacy

dimensions—process-metadata completeness, service-relevant fatigue/aging, independent validation, dataset comparability/traceability and external generalizability. Because the corpus mixes experiments, reviews and computational studies, these dimensions were used for qualitative theme-level appraisal rather than a single numerical risk-of-bias score.

Synthesis was causal rather than specialty-based. Findings were mapped from process or surface input to produced state and then to performance. Conflicting results were interpreted by examining differences in material system, machine/process metadata, post-processing, specimen geometry, aging, test method and model assumptions. Numerical pooling was avoided when datasets were not sufficiently homogeneous. Fig. 2 summarizes the selection and coding workflow; Table 2 provides the corresponding evidence-synthesis matrix.

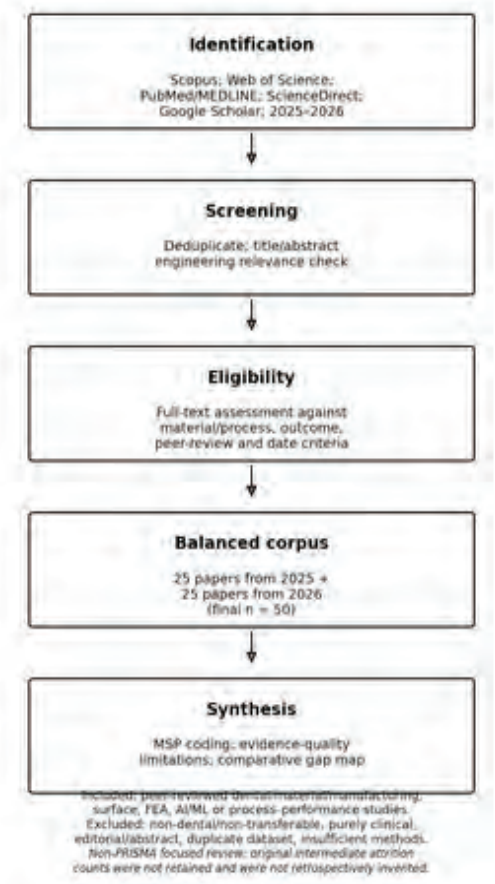


Fig. 2. Study-selection and evidence-synthesis workflow for the focused 2025–2026 critical review

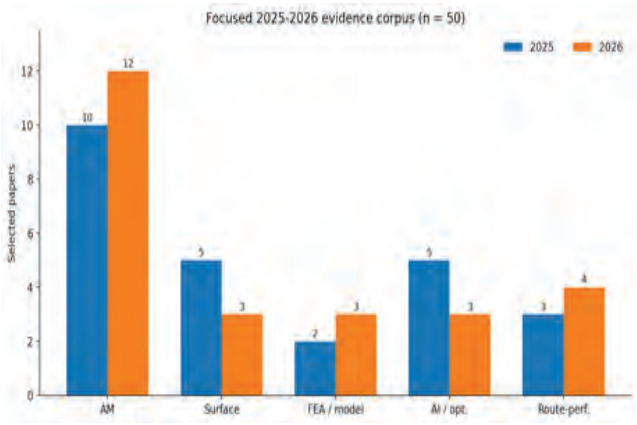


Fig. 3. Distribution of the focused 50-paper evidence corpus by engineering theme and publication year

Table 1. Engineering taxonomy of advanced dental material systems

Material system	Manufacturing state / risk	High-value engineering response
Zirconia / ceramics	Orientation, layer, sintering; pores, flaw direction, residual stress	Accuracy + Weibull/fatigue + fracture-origin mapping
Definitive / denture resins	Exposure/post-cure, conversion gradients, anisotropy, water ageing	Conversion + fit + fatigue/creep + ageing
PEEK / high-performance polymers	Raster/tool condition, crystallinity, fusion and surface energy	Wear + modulus + bonding + fatigue
Ti / Co-Cr alloys	PBF/machining; porosity, residual stress, roughness, passive film	Fatigue + tribocorrosion + dimensional accuracy
Surface-engineered systems	Laser/blasting/anodization/coatings; adhesion and oxide integrity	Adhesion + fatigue + wear/corrosion + biofunction
Patient-specific digital systems	Geometry, mesh, contact, load, material maps and uncertainty	Validated prediction + optimized design window

Table 2. Evidence-synthesis matrix for the 50 included publications

Theme / refs.	Critical synthesis / limitation
AM resins [3,4,21–27,45–47]	Orientation, post-cure and aging effects are consistent; resin/printer/protocol heterogeneity limits transfer.
Ceramic/zirconia AM [5,6,19,28–30,48]	Layer, orientation and firing affect accuracy/strength; defect-resolved fatigue remains sparse.
Metal/PEEK/routes [7,18,20,42]	Route alters surface integrity/durability; process metadata and coupled fatigue evidence are inconsistent.
Surface [8–11,31–34]	Multifunctional surfaces advance faster than adhesion and long-term wear/corrosion validation.
FEA [12,35–38]	Realistic contacts, patient-specific inputs, fatigue, uncertainty and independent validation remain incomplete.
AI/ML [13–15,39–41,43,49]	Speed gains are clear; small heterogeneous datasets and limited external testing restrict generalizability.

MANUFACTURING–STRUCTURE–PERFORMANCE ENGINEERING FRAMEWORK

Digital geometry and competing manufacturing routes

A manufacturing study begins before material is deposited or milled. Image segmentation, CAD

smoothing, mesh or tessellation tolerance, crown cement space, connector geometry, implant thread representation and support placement all influence the final geometry. These variables are often treated as software settings rather than manufacturing inputs, yet they determine wall thickness, local curvature, support accessibility and the direction of principal stresses relative to print layers. Shujaat and Jacobs [16] documented the large diversity of dental printers and technologies now reported in the evidence base, reinforcing the need to record machine and software details as traceable manufacturing metadata rather than generic labels such as '3D printed'.

For fixed dental prostheses, layer thickness and build orientation also change geometric fidelity through discretization and stair-stepping. Abdelhady et al. [17] synthesized evidence showing that layer thickness and orientation materially affect marginal and internal fit. The engineering lesson is broader than the specific numerical recommendations of any one review: dimensional accuracy is a response surface controlled jointly by layer, orientation, geometry and process capability. Therefore, a thesis that optimizes strength while ignoring fit can select an unusable process window; accuracy must be treated as a co-objective or constraint.

Additive and subtractive manufacturing do not simply produce different costs or lead times; they create different defect populations. Subtractive machining begins with a consolidated blank and introduces damage through tool contact, chipping, local heating, residual stress and edge geometry. Additive processes build the part from a liquid, slurry, filament or powder and introduce layer-related interfaces, pores, orientation effects and post-process distortion. Sari et al. [18] and Schettini et al. [7] provide useful route comparisons for implant-abutment and Co-Cr systems, while Pinelli et al. [19] and Ruggiero et al. [6] synthesize the printed-versus-milled zirconia evidence. The correct research question is therefore not 'Is AM better than milling?' but 'Under which process window does the AM defect population become comparable to or better than the damage introduced by the subtractive route for the required response?'

The same logic applies to CAD/CAM milling protocols. Pilecco et al. [20] showed that milling protocol can change fatigue behavior of restorative materials. Tool wear, feed, spindle speed, coolant, bur geometry and machining path should therefore be regarded as controlled variables in high-quality comparative research. For very brittle or thin components, edge damage and sub-surface flaws may dominate fatigue even when average surface roughness appears acceptable. Mechanical Engineering methods such as tool-life monitoring, acoustic emission, 3D surface metrology and residual stress measurement can bring needed rigor to dental machining studies.

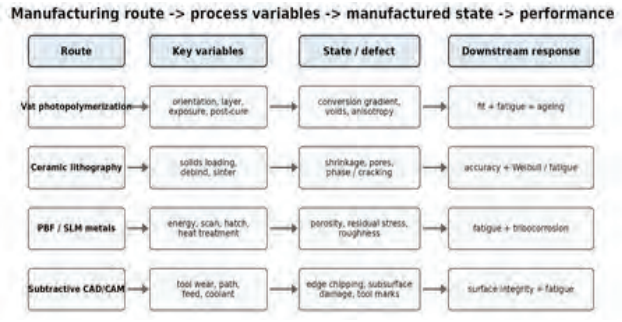


Fig. 4. Manufacturing-route/process-variable map linking additive and subtractive routes to characteristic defects and downstream performance.

Additive manufacturing of resins, ceramics, metals and high-performance polymers

Vat photopolymerization dominates many current dental AM applications because it can produce fine features with relatively high throughput. In DLP, SLA and LCD-type systems, final properties depend on optical exposure and chemistry as much as on geometric resolution. The printed 'green' network is typically incomplete and must be washed and post-cured. Polymerization continues during post-curing, increasing conversion and stiffness but also changing residual stress, dimensions and potentially brittleness. Rizzante et al. [3] and Korkmaz et al. [21] show why post-curing device, time and method should be recorded and optimized rather than treated as a fixed manufacturer instruction.

Build orientation changes the relation between the interlayer architecture and the applied stress. Mudhaffer et al. [4] reported orientation- and aging-dependent changes in flexural behavior of restorative resins,

while Alhotan et al. [22] and Terán Lozada et al. [23] demonstrate material-dependent physicomechanical performance of definitive printed systems. For Mechanical Engineering, orientation should be handled as an anisotropy problem. The research program should quantify whether the material can be represented by an orthotropic elastic/fatigue model, whether anisotropy decreases after post-curing, and whether crack paths align with layer boundaries or local porosity.

Denture-base resins add impact, water uptake, dimensional stability and surface/biological requirements. Coldea et al. [24] and Artopoulou et al. [25] show large material-to-material variability, and Abozaid et al. [26] emphasize that mechanical, optical, biological and clinical behavior should be considered together. High-impact formulations must also demonstrate retention of properties after thermomechanical aging rather than only higher initial strength [27]. A strong PhD therefore links formulation/process variables to network conversion, fracture surface, ageing kinetics and impact/fatigue response.

Ceramic AM is more complex because the printed geometry is not the final geometry. Slurry-based or lithography-based printing produces a green part that undergoes debinding and sintering; both steps introduce substantial shrinkage and can amplify density gradients, support effects and thermal distortion. Alghauli et al. [5] found a consistent influence of print orientation on the physico-mechanical behavior and accuracy of dental ceramics, and Ruggiero et al. [6] confirmed that build direction, technology and layer thickness affect the comparison between additive and subtractive zirconia. The manufacturing problem is therefore multiphysics: optical curing creates the green architecture, then thermal processing determines the final microstructure and residual defect state.

A defect-informed approach should measure density, pore size/orientation and surface flaw distribution in addition to flexural strength. Pinelli et al. [19] highlights the value of Weibull statistics when comparing printed and milled zirconia, while Sabatini et al. [28] links firing and build orientation to flexural and structural behavior. Bergler et al. [29] further indicates that mono- and multilayer zirconia should not be treated as equivalent because layer chemistry and microstructure may alter the response. For multilayer materials, the position of each layer relative to the tensile field becomes a design variable, motivating layer-resolved microhardness, phase, grain and fracture-origin mapping [30].

The scalar shrinkage value is not sufficient when shrinkage is anisotropic. A research-grade study should obtain Sx, Sy and Sz, quantify shape distortion with 3D scanning, and correlate the tensor-like shrinkage response with print orientation, support configuration, layer thickness, sintering schedule and local solid loading. Such a dataset can feed compensation algorithms that pre-deform CAD geometry to reduce final error.

Powder-bed fusion of Ti and Co-Cr enables customized frameworks, porous implants and complex geometries that are difficult to machine. However, the process introduces rough as-built surfaces, residual stress and internal pores that strongly affect fatigue. Schettini et al. [7] provides a direct manufacturing-route comparison for Co-Cr alloys. The next research step is to connect energy input and scan strategy to pore morphology, residual stress and fatigue using micro-CT, metallography and fracture mechanics rather than reporting only yield strength or hardness. Surface finishing should be treated as part of the final route because it can remove critical near-surface flaws while simultaneously altering dimensions.

PEEK and related polymers occupy a special position because their lower modulus can modify load transfer compared with metal or ceramic components. Their performance depends on both bulk processing and surface treatment for bonding. Sari et al. [18] compared additively/subtractively manufactured zirconia and PEEK implant abutments after thermomechanical aging, showing why manufacturing route and component connection must be considered together. A Mechanical Engineering contribution could combine polymer processing, crystallinity and surface treatment with wear/fatigue and validated contact mechanics.

$$S_L = [(L_0 - L_f)/L_0] \times 100\% \quad (1)$$

Table 3. Additive-manufacturing variables that should be controlled and reported

AM route	Mandatory process metadata	Recommended response set
SLA/DLP/LCD	Machine, wavelength, layer, orientation, exposure, support, wash, post-cure	Fit + conversion + anisotropic flexure/fatigue + ageing
Ceramic lithography	Solids loading, orientation, layer, debind, sinter, support	3D shrinkage + density + microstructure + Weibull/fatigue
FDM/FFF	Nozzle/bed temperature, raster, infill, speed, chamber/cooling	Density + DMA/DSC + orthotropic mechanics + wear/fatigue
PBF/SLM/DMLS	Power, scan speed, hatch, layer, build direction, heat treatment	micro-CT + residual stress + fatigue + tribocorrosion + accuracy
Hybrid AM + machining	AM history, stock allowance, toolpath and finishing	Surface integrity + dimensional accuracy + residual fatigue

RESULTS AND DISCUSSION

Process-structure-property-performance relationships

The performance of an additively manufactured part is often governed by the largest critically oriented defect rather than average porosity. A pore that is harmless in compression may become critical when placed near a tensile surface or connector. Therefore, density alone is an incomplete quality metric. Defect characterization should include size distribution, sphericity, orientation, distance from surface and relationship to the principal stress field. micro-CT provides a non-destructive route to build this descriptor and can be registered directly to a finite-element mesh to create defect-informed models.

Where Pf is probability of failure, sigma0 is characteristic strength and m is the Weibull modulus. The value of m should not be interpreted as a universal material constant because specimen geometry, surface finish and process state change the active flaw population. Pinelli et al. [19], Ruggiero et al. [6] and Bergler et al. [29] collectively motivate a process-specific reliability map in which orientation, layer and sintering are independent variables and m, sigma0, density and critical flaw size are responses.

In printed resins, the analogous hidden state variable is polymer network development. Degree of conversion, crosslink density, residual monomer and water uptake control stiffness, strength, creep, wear and biocompatibility. Because post-curing is spatially dependent on light penetration and geometry, a single surface measurement may not represent the interior. FTIR or Raman measurements at multiple depths, combined with microhardness or nanoindentation, can identify conversion gradients. Rizzante et al. [3] and Korkmaz et al. [21] show that post-curing conditions change final properties; the next step is to identify a process law that predicts properties from energy dose and thermal history.

where Y may be conversion, flexural strength, water sorption or fatigue life and Xi are process variables such as orientation, layer, exposure and post-cure time. The equation is not a substitute for mechanism; its value is to reveal interactions and identify regions for mechanistic follow-up. A good PhD uses DOE/RSM to reduce the design space, then uses microscopy, spectroscopy or FEA to explain why the optimum occurs.

Surface roughness is often reported as Ra or Sa, yet identical average roughness can conceal different sharpness, spacing, directionality and defect morphology. A polished sinusoidal-like surface and an air-abraded surface can have similar Ra but very different stress concentration, wettability and antagonist interaction. For advanced surface studies, 3D parameters such as Sa, Sq, Ssk, Sku, peak density and developed interfacial area are more informative. The surface should also be imaged and its functional response measured through friction, wear, contact angle, coating adhesion or corrosion rather than treated as a purely geometric descriptor.

This is particularly important for zirconia bonding treatments and implant coatings. Cantelli et al. [31] show that surface treatment influences fatigue survival of bonded zirconia, while da Silva et al. [9] demonstrates that an antimicrobial glass coating can change fatigue life and antagonist wear. These studies illustrate why surface engineering requires simultaneous evaluation of intended function and mechanical side effects.

$$P_f = 1 - \exp[-(\sigma/\sigma_0)\text{^}m] \quad (2)$$

$$Y = \beta_0 + \Sigma\beta_i X_i + \Sigma\beta_{ii} X_i^2 + \Sigma\beta_{ij} X_i X_j + \varepsilon \quad (3)$$

Surface engineering, accuracy and multifunctional interfaces

Research gap and contribution. Surface-engineering literature is rich in novel treatments but comparatively weak in coupled durability evidence. Initial wettability, antibacterial response or bond strength is frequently reported without demonstrating that the modified interface survives fatigue, wear, corrosion and aging.

The MSP framework therefore treats surface state as a manufactured interface whose functional benefit must be co-validated with adhesion and mechanical durability.

The engineering objective of a dental surface is not simply to make it rough or smooth. Implant surfaces must support initial stability, osseointegration, soft-tissue response and resistance to microbial colonization while retaining corrosion and fatigue integrity. Restorative surfaces must balance esthetics, antagonist wear, polishability, bonding and crack resistance. Surface engineering is therefore a multi-objective problem in which an intervention can improve one response while degrading another.

Recent reviews classify implant surface modification into subtractive treatments, physicochemical activation, laser processing, anodization and coating-based approaches. Aissi et al. [8], Abdo et al. [32], Cheng et al. [33], Jadhav et al. [11] and Balang et al. [10] collectively show a trend from conventional microroughening toward hierarchical micro/nano structures, bioactive chemistry, antibacterial function and local drug delivery. From an engineering perspective, every added layer creates an interface; therefore, coating adhesion, residual stress, crack resistance and stability under cyclic contact, sterilization and corrosion are essential translation metrics.

Laser surface engineering is attractive because texture can be created without adding a foreign coating, while process parameters can be digitally controlled. However, fluence, pulse duration, scanning overlap and thermal accumulation can change roughness, oxide chemistry and residual stress. A research program should identify a processing window that improves wetting or cell-relevant topography without inducing microcracking or fatigue reduction. Similar logic applies to blasting and acid etching: the treatment should be characterized by the resulting state variables rather than only by the nominal protocol.

Coatings introduce additional design dimensions: material selection, thickness, porosity, elastic modulus, adhesion, thermal expansion mismatch and dissolution rate. A thin hard coating can improve wear but may delaminate under bending; a bioactive layer can accelerate interface response but may be mechanically weak; an antimicrobial coating can lose function after wear. Pytel et al. [34] further adds manufacturing sustainability to this problem, highlighting energy and environmental indicators as future selection criteria.

Dental manufacturing differs from many engineering applications because a component can be mechanically strong yet clinically unusable if its fit is poor. Accuracy must therefore be incorporated into process optimization as a response or constraint. Printed fixed prostheses exhibit errors from pixel/voxel resolution, layer thickness, orientation, support, resin shrinkage and post-curing. Ceramic parts add sintering shrinkage; metal PBF adds thermal distortion; subtractive machining adds tool radius and access limitations. Abdelhady et al. [17] provides a current synthesis of layer-thickness and build-orientation effects on fit, while Shujaat and Jacobs [16] highlights the breadth of printer technologies and the need for evidence-based machine selection.

Dimensional verification should ideally use a full-field deviation map rather than a few marginal measurements. The process can be represented as: nominal CAD -> printed/milled part scan -> registration -> signed deviation field. This field can be decomposed into global scale error, warpage, local support distortion and surface stair-step. For ceramic AM, pre-sinter and post-sinter scans can isolate thermal shrinkage from printing error. Such data can be used to create machine/process-specific compensation factors or geometry morphing algorithms.

A high-value engineering objective is to optimize the Pareto trade-off among accuracy, strength and production time. Thinner layers may improve accuracy but increase build time; orientations that reduce supports may change mechanical anisotropy; aggressive post-processing may improve surface quality but remove dimensional stock. The result should be a process window rather than a single 'best' setting.

Table 4. Surface-engineering routes and validation requirements

Route	Primary mechanical risk	Minimum validation set
Polishing / grinding	Subsurface flaws; phase/residual-stress change	3D roughness + SEM + fatigue/wear
Air abrasion / etching	Sharp flaws, contamination, strength reduction	Surface damage + bond fatigue + fractography
Laser texturing	Thermal damage, microcracks, residual stress	Topography + chemistry + residual stress + fatigue
Anodization / activation	Brittle/porous oxide; altered corrosion	Oxide thickness + adhesion + electrochemistry + fatigue
Bioactive / smart coating	Delamination, mismatch, dissolution or wear loss	Adhesion + cyclic shear + corrosion/ageing + function/safety

Predictive finite-element mechanics and validation

Research gap and contribution. Dental FEA has progressed faster in geometric sophistication than in parameter and validation fidelity. The key gap is weak transfer of measured manufacturing state—anisotropy, porosity, residual stress, surface/interface behavior and as-built geometry—into model inputs. The framework therefore defines model fidelity through measurable state variables, sensitivity analysis and independent experimental validation rather than visual complexity.

FEA is most valuable when it tests a defined mechanical hypothesis. For example, if a new implant thread is intended to reduce cortical stress, the model should isolate geometry while maintaining comparable material and contact assumptions. If a printed polymer framework is proposed to alter load transfer, measured anisotropic properties should be used. Poovarodom et al. [35] and Sahin Hazir et al. [36] demonstrate how material choices change stress distribution, but broader review evidence indicates that many dental FEA studies still rely on isotropic linear properties, perfect interfaces and static loads [12, 37].

Model fidelity can be organized as a ladder rather than an all-or-nothing choice. Level 1 uses simplified geometry and linear elastic static analysis to understand trends. Level 2 introduces anatomical geometry, mesh convergence and realistic contact. Level 3 adds preload, friction, nonlinearities and oblique/dynamic load. Level 4 adds fatigue, remodeling, patient-specific material maps and uncertainty. Level 5 combines experimental validation with surrogate or digital-twin functionality. Fig. 5 emphasizes that higher complexity is only meaningful when parameters can be justified or measured.

Implant systems contain multiple contacts: crown-abutment, abutment-implant, screw threads and bone-implant interface. Treating all contacts as perfectly bonded eliminates slip and can underestimate local stress concentration. Screw preload changes the normal force and frictional capacity of the joint. Similarly, osseointegration is not binary during healing; interfacial stiffness evolves with time. A model intended to predict mechanical complications should therefore justify contact, friction and preload, perform sensitivity analysis, and report how results change when interface assumptions are varied.

Padullés-Roig et al. [38] is particularly instructive because fatigue testing and FEA are combined around a clinically motivated geometry change. This experiment-simulation combination is preferable to isolated FEA because the model can be challenged against component behavior. A validation metric can include strain-field error, displacement, stiffness, load-to-failure or failure location. Agreement in a single scalar stress is not validation because stress is not directly measured in most laboratory setups.

The next generation of dental FEA should include manufacturing state variables. For printed zirconia or metals, micro-CT-derived pores can be imported or represented statistically. For printed polymers, orthotropic properties can be assigned by build direction. For multilayer zirconia, layer-specific elastic/fracture properties can be mapped spatially. For surface coatings, cohesive-zone interfaces can simulate delamination. These approaches convert FEA from a geometry-only tool into a process-aware model.

where K depends on displacement u and a set of material, contact and process-state variables X. In linear models K is constant; in nonlinear contact, damage or material models it evolves. The research challenge is to identify which components of X materially affect the clinical/mechanical response and then measure them. This sensitivity-driven approach prevents unnecessary model complexity while improving predictive value.

$$[K(u, X)]\{u\} = \{F\} \quad (4)$$

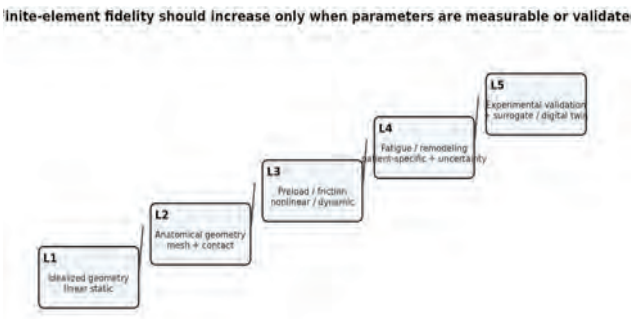


Fig. 5. Finite-element fidelity ladder from idealized screening analysis to validated patient-specific digital twins.

Table 5. Minimum reporting and validation checklist for dental finite-element studies

Domain	Minimum requirement	Preferred advanced practice
Geometry / mesh	Source, simplification, element type, size, convergence	Patient-specific scan + dimensional verification + adaptive refinement
Materials	E, nu and source	Measured anisotropic/nonlinear/viscoelastic process-specific data
Contact	Bonded/frictional definition	Measured friction, preload and interface law + sensitivity
Loading	Magnitude, direction and location	Variable/dynamic spectrum and muscle/occlusal constraints
Validation	Comparison with independent data	DIC/strain/displacement/failure-mode validation
Uncertainty	At least sensitivity analysis	Monte Carlo/probabilistic inputs and confidence bounds

Machine learning, surrogate models and digital twins

Research gap and contribution. AI/ML applications demonstrate speed and optimization potential, but generalizability is limited by small datasets, inconsistent metadata, simulation-generated labels and narrow material/machine domains. The proposed contribution is a traceable physics–data workflow in which domain coverage, external testing, uncertainty and high-fidelity confirmation are mandatory before optimization claims are accepted.

Design of experiments remains the most appropriate starting point when data are limited and factors are controllable. Full/fractional factorial designs, Taguchi arrays, central-composite designs and Box-Behnken designs can identify main effects and interactions. RSM provides interpretable polynomial models, but high-dimensional nonlinear relationships may require random forests, gradient boosting, Gaussian processes or neural networks. The key transition is from a statistical fit to a validated surrogate capable of predicting unseen combinations within a defined domain.

In manufacturing, surrogate modelling is valuable because both experiments and high-fidelity FEA are expensive. Mohammadnabi et al. [39], Shangguan et al. [40] and Soori et al. [41] describe AI applications across material development, parameter optimization, monitoring and post-processing. Dental research can use the same architecture: generate a structured dataset from DOE-guided experiments and calibrated simulations, train a surrogate, quantify its error and uncertainty, then use the surrogate in an optimizer. The original experiment/FEA must remain the ground truth for final validation.

Dental implant design provides one of the clearest current demonstrations of this approach. Chen et al. [13] generated a large FEA dataset, trained an artificial neural network and used particle swarm optimization to identify improved complete-arch implant configurations with near-real-time computation. Prados-Privado [14] extended the concept toward numerical inverse design using neural surrogates and evolutionary optimization. These studies suggest a future in which patient geometry and bone quality become inputs to an optimization engine rather than being assessed through a small set of manually selected implant designs.

The attraction of surrogate models is computational speed, but three controls are essential: (1) the training domain must cover the intended design space; (2) an independent test set must quantify generalization; and (3) optimized solutions must be re-evaluated with high-fidelity FEA and preferably experiment. Extrapolation beyond the training domain should be explicitly rejected or associated with high uncertainty.

A digital twin differs from a static FEA model because it is intended to update as the physical system or process changes. In dental implants, healing changes bone-implant contact and interfacial stiffness. Rodriguez-Molinero and Prados-Privado [15] combined FEA-derived states, osseointegration dynamics and deep learning to produce time-resolved biomechanical predictions. The study is a proof-of-concept and still requires patient-specific biological validation, but it establishes a useful research architecture: physics-based simulation generates interpretable states, mechanobiology updates them over time, and a surrogate reduces computation.

A manufacturing digital twin can be constructed similarly. Sensor data from an AM machine (exposure, temperature, layer images, energy, recoater signals) can update a process model; metrology and mechanical tests can update the property model; and optimization can adjust future builds. For dental manufacturing, the biggest barrier is not algorithm availability but standardized, traceable datasets that connect machine settings to final component performance.

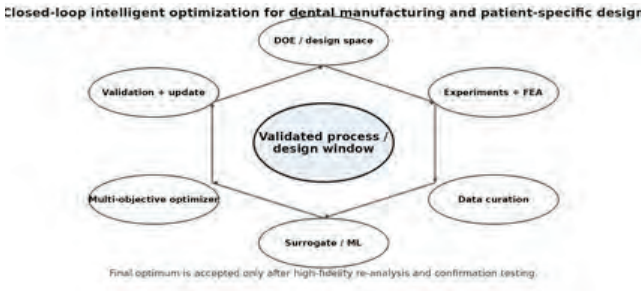


Fig. 6. Closed-loop intelligent optimization architecture for dental manufacturing and patient-specific component design.

4.5 Multi-objective process/design optimization

Most dental-material studies optimize a single response, yet practical design is inherently multi-objective. A zirconia process should maximize reliability and accuracy while minimizing roughness, shrinkage error and production time. A printed resin process must balance strength, fatigue, water sorption, fit, conversion and optical stability. An implant surface should balance osseointegration, antibacterial function, coating adhesion, fatigue and tribocorrosion. A patient-specific implant design must reduce bone stress without creating excessive implant stress or violating surgical constraints.

where x contains material, process, geometry and surface variables. The result is normally a Pareto set rather than one mathematically unique solution. Methods such as NSGA-II, particle swarm optimization, differential evolution, Bayesian optimization and desirability functions can be used depending on the number of variables, computational cost and uncertainty. The objective should not be to use a fashionable optimizer; it should be to produce a physically interpretable process/design window that can be experimentally validated.

Robust optimization is especially important because dental materials exhibit scatter and patient conditions vary. A nominal optimum with extremely steep sensitivity may be undesirable in manufacturing. Process robustness can be quantified by variance, probability of failure or worst-case response across plausible parameter distributions. This transforms a research contribution from 'optimal setting at laboratory conditions' into 'manufacturing window that retains performance under realistic process variability'.

Sustainability can be added as another objective. AM may reduce stock waste but increase photopolymer, support, cleaning and post-curing requirements; surface treatments may consume energy and chemicals; ceramic sintering is energy intensive; milling consumes tools and blanks. Pyteľ et al. [34] identifies the shortage of standardized sustainability indicators in dental surface engineering. A future mechanical-engineering research program can combine quality, fatigue, energy, waste and cost in a lifecycle-aware Pareto analysis.

$$\min F(x) = [f_1(x), f_2(x), \ldots, f_k(x)], \text{ subject to } g_j(x) \leq 0 \text{ and } h_l(x) = 0 \quad (5)$$

Advanced composites, functionally graded structures and architected lattices extend the design space beyond homogeneous material selection. The source review emphasizes that filler architecture, multimaterial interfaces, designed porosity and spatial gradients should be measured in the as-manufactured state, then linked to anisotropic constitutive behaviour, fatigue and failure location. Sustainability and process economics should be treated as quantified co-objectives rather than assumed advantages of additive manufacturing; measurable inputs include material waste, energy, post-processing, tool consumption, rejected components and validated service performance [40, 41].

Advanced composites and hybrid dental materials

Advanced dental composites should be treated as hierarchical engineered systems rather than mixtures described only by matrix and filler percentages. Polymer chemistry controls conversion, viscoelasticity and water uptake; filler morphology governs stiffness, crack deflection and polishing response; interfaces determine load transfer and hydrolytic stability; and the manufacturing route controls filler orientation, voids, cure gradients and surface texture. Recent 3D-printing literature illustrates this coupling because nominally similar photopolymer systems can exhibit different flexural response, aging sensitivity and dimensional behavior after changes in orientation and post-curing [4, 3, 23]. A Mechanical Engineering review should therefore describe a composite by a state vector that includes composition, filler architecture, conversion, porosity, residual stress and surface condition, not by commercial class alone.

For particulate and nanofilled systems, the key design issue is the competition between reinforcement and defect creation. Increasing filler fraction can raise modulus and hardness but can also increase viscosity, limit printability, promote agglomeration and create stress concentrations if interfacial bonding is poor. For short- or continuous-fibre systems, fibre length, orientation, volume fraction and interface quality produce strong anisotropy that can be beneficial when aligned with functional load paths and harmful when the clinical load rotates relative to the build direction. This suggests a PhD pathway based on processing-induced architecture: quantify local reinforcement orientation by microscopy or micro-CT, derive orthotropic constitutive properties, and validate component-level FEA under axial and oblique loading.

Hybrid material systems also create interfaces between materials with different elastic modulus, thermal expansion, fracture behavior and water sorption. Such interfaces can redistribute stress or become preferential crack paths. The correct research question is therefore not whether a hybrid material is 'stronger', but whether the interface remains stable under combined thermal, moisture and cyclic mechanical exposure. A useful experimental program can combine lap/shear or interfacial fracture tests, thermocycling, cyclic loading and fractography, then use cohesive-zone modelling to convert observed interface degradation into predictive component simulations.

Machine learning is attractive for composition development because the variable space can include resin chemistry, filler type, particle size, silanization, fraction, print settings and cure schedule. However, data-centric material discovery requires disciplined metadata and physically meaningful descriptors. Shangguan et al. [40] and Mohammadnabi et al. [39] emphasize the broader potential of AI in additive material/process development; for dental materials, the next step is to create compact but high-quality datasets where each formulation is linked to rheology or printability, conversion, defect state, mechanical response and aging. Physics-informed descriptors such as filler surface area, modulus contrast, cure dose and porosity are more transferable than one-hot commercial-material labels.

Functionally graded and multimaterial structures

Natural teeth are functionally graded: enamel, dentin and the dentino-enamel junction create spatial variations in modulus, toughness and crack resistance. Most manufactured restorations and implants remain comparatively homogeneous, producing abrupt stiffness transitions at cement, abutment, bone or veneering interfaces. Functionally graded materials (FGMs) offer a Mechanical Engineering strategy in which composition, porosity, phase fraction or architecture varies continuously or stepwise with position. The design target is not simply a high bulk strength; it is a controlled spatial property field E(x), K_IC(x), hardness H(x) or porosity p(x) that manages contact stress, protects interfaces and directs damage away from critical regions.

In a crown, a graded ceramic/polymer or ceramic-composition architecture could combine a hard wear-resistant occlusal surface with a more compliant interior that reduces tensile stress at the cement interface. In an implant, a dense high-strength core may transition to a porous or bioactive outer zone to reduce effective stiffness and improve bone ingrowth. In a framework, topology and material grading can be coupled so that local stiffness is concentrated along load paths while non-critical volume is reduced. These concepts align naturally with additive manufacturing, which can control geometry layer-by-layer and, in selected platforms, vary material or energy input spatially.

The main research barriers are manufacturability and verification. A designed gradient is not necessarily the gradient that is produced after photopolymerization, debinding, sintering, thermal shrinkage or multi-material bonding. Therefore, a defensible PhD should distinguish intended digital gradient, manufactured gradient and effective mechanical gradient. Micro-CT, microscopy, spectroscopy or nanoindentation maps can characterize the actual spatial state; FEA should then use the measured rather than nominal property distribution. The resulting model can answer whether a gradient genuinely reduces critical stress or simply introduces a new weak transition zone.

Optimization of FGMs is intrinsically multi-objective. Increasing compliance may reduce stress shielding but increase micromotion; increasing surface porosity may

support tissue integration but reduce fatigue strength; adding a tough polymer-rich zone may improve damage tolerance but reduce wear resistance. The design variables can be represented by a small number of gradient-control parameters and optimized using a surrogate-based framework. Validation should include not only peak load but cyclic survival and failure origin, because a graded design is successful only when it changes the mechanism and location of damage in a favorable way.

Porous, lattice and architected implants

Architected materials extend dental design beyond conventional solid geometry. Lattice cells, stochastic porosity and triply periodic minimal surface (TPMS) structures can reduce mass, tune effective elastic modulus and create interconnected pore networks while retaining a load-carrying skeleton. For metallic implants, the principal motivation is stiffness matching and biological integration; for polymeric or hybrid prostheses, lattices can also control compliance, damping and material use. The engineering design space includes unit-cell topology, relative density, strut or wall thickness, pore size, orientation, grading and local reinforcement.

A critical distinction must be made between geometric porosity that is intentionally designed and manufacturing porosity that is an uncontrolled defect. Designed pores can improve functional architecture, whereas lack-of-fusion pores, unmelted particles, trapped resin or ceramic voids can initiate fatigue cracks. Micro-CT is therefore particularly valuable: it can separate intended geometry from defects, quantify strut deviations and provide a mesh for as-manufactured FEA. This closes the process-structure-performance chain and avoids the common assumption that the CAD lattice is identical to the manufactured lattice.

The fatigue problem is central. Lattice structures contain many stress raisers, and their behavior depends on surface roughness, residual stress and build orientation. A static compression test can screen stiffness but cannot establish long-term implant safety. A research study should combine coupon-level cell testing, component-level cyclic loading, fractography and probabilistic modelling. Surface finishing should be included because it can change both roughness and strut dimensions. For coated porous structures, coating penetration, thickness uniformity and local adhesion should be characterized rather than inferred from an external flat specimen.

Computationally, architected structures create an opportunity for multiscale modelling. Homogenized properties can make whole-implant optimization computationally practical, while local submodels resolve critical cells near threads or interfaces. Surrogate models can then explore large topology/relative-density spaces. The optimized design must be checked against manufacturability constraints such as minimum wall thickness, trapped powder/resin removal, support accessibility and post-processing. Thus the true objective is a manufacturable architecture with controlled variability, not an ideal mathematical lattice.

Sustainability and lifecycle-aware optimization

Sustainability should enter dental manufacturing as a measured engineering response rather than a qualitative claim that additive manufacturing is automatically ‘green’. Subtractive manufacturing generates machining chips, partially used blanks and tool wear but often has mature workflows and predictable productivity. Additive manufacturing can reduce geometric material waste, yet it may require supports, failed builds, disposable vats or films, solvent washing, inert gas, post-curing, debinding and high-temperature sintering. The net outcome depends on material class, batch size, equipment utilization and the functional lifetime of the restoration or implant.

A useful lifecycle boundary for a PhD can begin at digital manufacturing and end at a validated component, thereby avoiding unsupported whole-clinic assumptions. Measurable inputs include mass of virgin and discarded material, solvent use, support mass, machine time, electrical energy, furnace energy, tool consumption, post-processing consumables and rejected components. These can be normalized per accepted component and then coupled with quality responses such as fit, strength and fatigue life. A process with low energy per build may be less sustainable if its rejection rate or replacement frequency is high.

Process economics can be analysed in parallel. Cost per accepted part can be decomposed into machine

amortization, material, labor, energy, tooling, post-processing, inspection and scrap. Sensitivity analysis reveals which assumptions dominate. For small-batch patient-specific production, digital flexibility may outweigh raw material efficiency; for standardized high-volume components, throughput and post-processing may dominate. The analysis should therefore report the production scenario explicitly rather than producing a universal ranking.

The most useful final output is a multi-objective Pareto map involving performance, manufacturing time, cost and environmental burden. Pytel' et al. [34] notes the importance of smart and sustainable surface-engineering strategies, while AI/AM reviews describe the possibility of data-driven process control [41, 39]. Combining these themes creates a new doctoral direction: a closed-loop manufacturing system that rejects low-quality states early, reduces unnecessary reprocessing and optimizes a verified performance-to-resource ratio.

Evidence synthesis and research maturity

The focused corpus shows a consistent maturity gradient. AM research is comparatively mature for static mechanical and dimensional characterization but less mature for defect-resolved fatigue, cross-machine transfer and long-term aging. Surface-engineering research increasingly integrates topography, chemistry and biofunction, yet coating adhesion and retained function after cyclic wear/corrosion are less consistently demonstrated. FEA is mature as a comparative hypothesis tool but less mature as a predictive tool when patient-specific properties, realistic contacts, fatigue, uncertainty and experimental validation are absent. AI/ML shows rapid computational progress but the weakest evidence for external generalization because datasets are often small, heterogeneous or simulation-derived.

Evidence quality must therefore be interpreted by the completeness of the causal chain rather than by novelty alone. A statistically significant increase in static strength does not establish fatigue reliability; an antimicrobial coating is not durable unless function is retained after mechanical/environmental exposure; a detailed stress map is not validation; and low internal ML error does not demonstrate transfer to an unseen printer, material or geometry. Table 6 makes these limitations and minimum strengthening actions explicit.

The strongest future studies should connect process parameter → measured state → service-relevant degradation → validated model → uncertainty-aware optimization. Fig. 7 summarizes the qualitative maturity pattern used to derive the translational roadmap in Section 5. The categories are comparative descriptors within this focused corpus and not formal risk-of-bias scores.

Table 6. Evidence-quality limitations and minimum strengthening actions

Theme	Evidence limitation and minimum strengthening action
AM	Heterogeneous machines/materials and static-test dominance; require full metadata, cross-build replication, micro-CT/fractography and aging/fatigue.
Surface	Short-term functional endpoints and underreported durability; co-test adhesion, cyclic wear/tribocorrosion, fatigue and retained function.
FEA	Idealized properties/contacts and inconsistent validation; use measured inputs, sensitivity/ convergence and DIC/bench validation.
AI/ML	Small heterogeneous data and external-domain risk; report provenance, independent test performance, uncertainty and re-FEA/ experimental confirmation.
Cross-study	Non-standard metadata/endpoints; adopt common MSP state variables and minimum reporting sets.

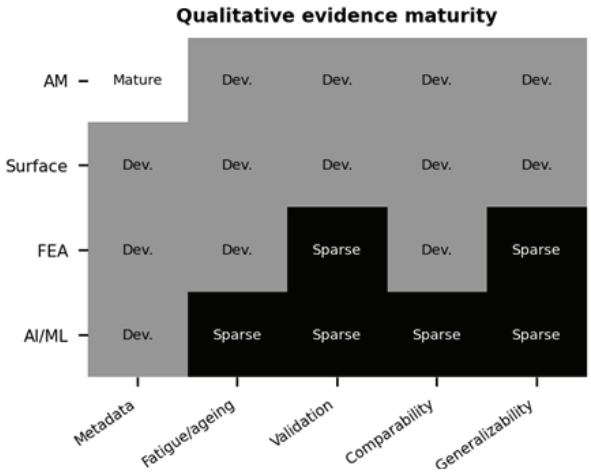


Fig. 7. Qualitative evidence-maturity map for the focused 2025–2026 corpus

RESEARCH GAPS, TRANSLATIONAL ROADMAP AND REPORTING PRIORITIES

A recurring limitation across dental manufacturing is that nominally identical variables do not always describe identical physical states. A label such as 50 µm layer thickness does not specify exposure history, resin age, build temperature, washing efficiency, post-cure irradiance or support-induced distortion. Likewise, a zirconia sintering temperature does not fully describe heating rate, dwell, cooling history, green density or furnace loading. Reproducibility therefore requires measurement of the state produced by the process. For polymers this may include degree of conversion, glass-transition behavior and water uptake; for ceramics, density, phase fraction, grain/defect population and shrinkage; for metals, porosity, residual stress and surface state.

Measurement uncertainty should be separated from material/process variability. A 3D scanner introduces registration and resolution uncertainty; micro-CT thresholding changes measured pore volume; surface roughness depends on filtering and scan area; fatigue life is intrinsically scattered. A research dataset should include repeated specimens, repeated measurements where appropriate and a measurement-system check before optimization. Otherwise, a machine-learning algorithm may model metrology noise as if it were a process mechanism. Statistical reporting should therefore include dispersion and confidence intervals, not only means and p-values.

For optimization, uncertainty should enter the decision variable explicitly. If two process settings have similar mean strength but one has a much larger variance, the robust setting may be preferable. A probability-of-failure criterion, lower confidence bound or tolerance-based objective can be used in addition to mean response. This is especially important for brittle zirconia, where Weibull-type reliability is more informative than an isolated average flexural strength, and for fatigue, where run-outs and censored observations should be retained rather than discarded [19, 6].

Benchmark-quality data also requires traceable metadata. Each specimen should be associated with material batch, machine, build job, position on platform, geometry file/version, process parameters, post-processing, operator-relevant steps, storage, aging, metrology and test condition. The resulting data object becomes useful for both conventional statistics and future ML. This is one reason a carefully designed PhD dataset of hundreds of traceable specimens can be more scientifically valuable than thousands of values extracted from heterogeneous publications.

Dental components do not experience one isolated degradation mode. Polymers absorb water while they are cyclically loaded; ceramics experience contact and slow crack growth in a humid environment; implants undergo micromotion, fretting and electrochemical reactions; coatings experience contact, corrosion and interfacial shear. Yet many laboratory studies isolate a single variable because it simplifies testing. The next research step is not to reproduce the entire mouth indiscriminately, but to identify the dominant coupled damage sequence for the application and reproduce that sequence with controlled boundary conditions.

For printed restorative resins, a useful hierarchy is dry static screening, water-conditioned properties, thermocycling, cyclic loading and then combined thermomechanical exposure. Mudhaffer et al. [4], Alhotan et al. [22] and Alhotan et al. [27] show why aging state matters when interpreting printed polymer performance. The research question should ask which microstructural or network variable changes during aging and whether this variable explains the change in fatigue or fracture. FTIR/Raman, DMA, sorption and fractography can therefore be paired with mechanical testing rather than treating aging as a black-box conditioning step.

For metallic implant surfaces and coatings, electrochemical behavior should be evaluated before, during or after mechanically assisted surface disruption depending on the hypothesis. Tribocorrosion testing can partition material loss conceptually into mechanical, electrochemical and synergistic contributions, while post-test SEM/EDS identifies oxide disruption, debris and coating delamination. The study should then determine whether the damaged state alters fatigue strength or interface behavior. This creates a more clinically meaningful chain than reporting a corrosion curve and a wear scar in unrelated specimens.

Accelerated ageing must also preserve mechanism. Raising temperature or load can shorten test time, but an excessively severe condition may activate a failure mode that is not dominant in service. A sound PhD therefore uses acceleration only after identifying the expected mechanism and includes at least one service-relevant confirmation condition. The strongest output is a validated acceleration rule or damage-equivalence criterion rather than simply a high number of cycles.

A recurring disconnect in dental materials research is the jump from simple coupon tests to claims about clinical components. Coupon specimens are essential for isolating material/process effects, but they remove geometric stress concentrations, cement layers, screw preload, occlusal contact, manufacturing supports and anatomical variability. Conversely, testing only complete crowns or implants makes it difficult to identify which mechanism caused the observed difference. A staged translation ladder resolves this tension.

Stage I should use standard or simple specimens to screen process variables and estimate constitutive, fracture, wear or fatigue parameters. Stage II should use a representative component geometry manufactured by the same process and surface route. Stage III should use experimentally validated FEA to explain local stress/strain and translate measured material behavior into component performance. Stage IV can introduce patient-specific geometry, bone quality or occlusal variability. Stage V, where feasible, should connect the engineering endpoint with dental/biological or clinical evidence through interdisciplinary collaboration. Each stage should have a predefined gate; failure to reproduce the mechanism at a higher stage should trigger model revision rather than selective reporting.

Manufacturing readiness can be assessed similarly. A process demonstrated on one printer and one batch is an exploratory result. Repeating the optimized setting across build positions, days or material lots tests robustness. Repeating on a second machine or laboratory tests transferability. Monitoring build signals and applying statistical process control tests whether the process is controllable. Only after these steps should an ML model be described as deployable or a process window as generalizable. Shujaat and Jacobs [16] emphasize the practical importance of evidence-based printer selection, while Chen and Wei [1] show how dental 3D printing spans very different material/process classes; both reinforce the need to define the domain in which a conclusion is valid.

This ladder also creates a publication strategy. The coupon/process paper can establish mechanism; the component/FEA paper can establish transferability; the optimization paper can establish design value; and the final translational study can establish practical relevance. Such a sequence is more coherent than four unrelated material-comparison papers because each paper reduces a specific uncertainty in the same causal chain.

Digital twins and cross-laboratory AI depend on interoperability. At present, manufacturing data may reside in printer software, geometry in STL/3MF files, metrology in proprietary scan formats, microscopy as images and mechanical results in spreadsheets. A future-ready research program should create a specimen-level identifier that links these sources without losing provenance. A lightweight data architecture can store immutable raw files, processed features, analysis scripts and versioned models. This is not merely data management; it permits a later researcher to determine exactly which geometry and process history produced a given fatigue result.

A minimum digital record should distinguish commanded process parameters from measured process signals and final-part characteristics. For example, commanded exposure time is not the same as measured irradiance; nominal strut diameter is not as-built diameter; nominal implant preload is not necessarily retained preload after settling. Digital-twin credibility depends on updating the model with measured states rather than assuming perfect execution of the digital plan. Sensor data, metrology and inspection therefore form the observational layer of the twin.

Interoperable datasets should also contain failure information. Fractography images, crack-origin labels, pore descriptors, wear maps and coating-delamination measurements can become multimodal data for interpretable ML. The scientific opportunity is to train models that predict not only a scalar strength value but a failure mode or spatial damage field. This is more aligned with Mechanical Engineering because it preserves the link between data and mechanism.

For researchers with limited access to advanced sensors, the same principle can be implemented at modest scale: standardized specimen IDs, machine logs, 3D scan deviations, SEM images and mechanical curves stored in a structured directory and linked by a metadata table. The research value comes from traceability and consistency, not from the size of the database. This architecture can later support transfer learning, federated studies or collaborative benchmark datasets across dental and Mechanical Engineering laboratories.

Table 7. Research-gap-to-method roadmap for a mechanical-engineering research program

Priority gap	Strong method chain	Validated deliverable
AM zirconia defects	DOE -> 3D metrology/micro-CT -> fatigue -> fractography -> FEA	Process-defect-fatigue reliability map
Printed resin cure engineering	Orientation/post-cure -> FTIR/Raman/DMA -> ageing -> fatigue	Cure-dose-conversion-fatigue window
Accuracy-strength-productivity	3D deviation + mechanics + build time -> surrogate/Pareto search	Robust clinically constrained AM process window
Smart coating durability	Coating state -> adhesion -> wear/tribocorrosion -> residual fatigue	Durability-function Pareto map
Manufacturing-informed FEA	As-built geometry/properties -> nonlinear contact FEA -> DIC/strain	Quantified model-fidelity benefit and validation protocol
Patient-specific inverse design	Space-filling FEA -> surrogate -> MOO -> re-FEA/bench confirmation	Uncertainty-aware inverse-design workflow
Digital twin interoperability	Traceable process + metrology + mechanics + time-state update	Time-resolved risk map with defined domain of validity
Sustainable intelligent manufacturing	Energy/material/cost + quality/life -> robust Pareto optimization	Performance-cost-environment decision map

Experimental, metrological and statistical discipline

A strong doctoral program should be built as an evidence chain rather than a collection of independent tests. If the hypothesis is that build orientation changes fatigue because it changes flaw orientation, the minimum chain is orientation -> defect/anisotropy measurement -> static screening -> fatigue -> fracture origin -> model validation. If the hypothesis is that a coating improves implant performance, the chain is deposition parameters -> thickness/roughness/adhesion -> wear/tribocorrosion or cyclic shear -> residual fatigue -> biological or functional endpoint. Each measurement should either quantify an input, a state variable or a validation response.

Standardized simple specimens and anatomically realistic components should be used at different stages. Bars/discs are useful for process screening because stress fields are interpretable; crowns, frameworks and implants are useful for final transferability. A common error is to begin immediately with complex anatomical parts, where material, geometry and contact effects are confounded. Conversely, a PhD that never progresses beyond bars may not demonstrate component relevance.

The statistical architecture should be planned at the same time as the mechanical experiment. Screening studies may use factorial or fractional-factorial designs to identify the dominant manufacturing variables, whereas optimization studies require response-surface or sequential designs that resolve curvature and interactions. Replication is needed to estimate process scatter; randomization should be used where practical to avoid confounding build order, furnace batch or testing sequence with the intended factors. For additive manufacturing, build position and build job should be recorded because specimens from one job are not always statistically independent. A mixed-effects formulation can distinguish fixed process factors from random batch/build effects when sufficient data are available.

Sample size should be justified by the expected effect and variability rather than copied from previous papers. Static mechanical tests may need fewer specimens than fatigue or Weibull reliability studies, because lifetime distributions are broader and run-outs contain information. Pilot data can estimate variance and inform power calculations, but the pilot settings should not be used repeatedly to tune the hypothesis and then treated as independent confirmation. For machine-learning studies, the effective sample size is the number of independent manufacturing conditions or components, not simply the number of rows created by repeated measurements. Splitting multiple observations from the same specimen across training and test sets creates data leakage and unrealistically high prediction accuracy.

Validation criteria should be defined before model fitting. For FEA, agreement can be expressed as error in strain, displacement, contact force, fracture load or failure location under one or more conditions not used in calibration. For a surrogate or ML model, report held-out error, calibration/uncertainty where applicable and performance on an unseen geometry, material batch or process setting. For optimization, confirmation experiments should include the predicted optimum plus at least one baseline/control, with the same measurement protocol. The relevant question is not

whether the optimizer returns a solution, but whether that solution produces a statistically and mechanically meaningful improvement when built and tested.

Finally, significance and engineering importance should be separated. A small change may be statistically significant yet irrelevant to fit tolerance or fatigue survival, while a clinically meaningful reduction in failure probability may not reach significance in an underpowered experiment. Reporting effect size, confidence intervals and physically defined acceptance thresholds allows the reader to distinguish these cases. This decision-oriented statistical framework is essential if a PhD aims to deliver a process window or design rule rather than a sequence of pairwise comparisons.

High-value research pathways

A doctoral problem becomes clear when five elements are specified: material, manufacturing/surface route, controllable variables, performance responses and method/validation. Figure 8 demonstrates this objective generator. The framework prevents titles such as 'Study of 3D-printed dental materials' because it forces the researcher to define what is controlled and what causal mechanism is being tested.

The strongest objectives typically use verbs that imply engineering knowledge: quantify, model, correlate, identify, predict, validate, optimize and establish a process/design window. 'Compare' can be one step but should not be the thesis endpoint. A publishable PhD should generate a relation or model that allows a result to be predicted outside the exact tested combination, followed by independent validation.

Quantify the effect of build orientation, layer thickness and sintering schedule on pore architecture, shrinkage anisotropy and fatigue reliability of additively manufactured zirconia.

Develop a cure-conversion-property model for definitive DLP/SLA crown resins linking post-curing energy and geometry to conversion gradients, water uptake, anisotropy and fatigue life.

Optimize printing orientation, layer thickness and support strategy simultaneously for marginal fit, flexural/fatigue performance and manufacturing time using a multi-objective DOE/ML framework.

Compare additive and subtractive routes through quantified defect populations rather than only mean strength, and develop a defect-informed reliability model for matched dental geometries.

Develop and validate a tribocorrosion-fatigue protocol for surface-engineered titanium implant materials that separates mechanical wear, electrochemical loss and synergistic damage.

Engineer a multifunctional implant coating and determine the trade-off among roughness, wettability, coating adhesion, cyclic shear/fatigue, corrosion resistance and antibacterial/bioactive function.

Develop manufacturing-informed FEA using measured anisotropic material properties, micro-CT defect maps and realistic contact; validate with DIC or strain-gauge measurements.

Create a patient-specific implant inverse-design framework using FEA-generated datasets, a calibrated surrogate model and constrained multi-objective optimization with independent re-FEA and experimental verification.

Develop a time-resolved digital-twin concept for implant healing that updates interfacial stiffness and loading risk using mechanobiological state variables and validated surrogate prediction.

Use in-situ or layer-image data from dental AM to detect defects and predict final dimensional/mechanical quality using interpretable machine learning.

Construct an open, traceable dental AM benchmark dataset containing machine metadata, process parameters, metrology, microstructure and mechanical responses to enable cross-laboratory ML validation.

Perform sustainability-aware optimization of dental manufacturing routes by including energy, waste, cycle time, cost and durability in a common Pareto framework.

Representative project blueprints

Blueprint A - Defect-informed AM zirconia: Hypothesis: the fatigue reliability of printed zirconia is controlled by orientation-dependent pore/flaw architecture created jointly by printing and sintering. Use a factorial orientation-layer-sintering design, 3D scan/micro-CT,

density/XRD/SEM, flexure/Weibull and step-stress fatigue. Register representative defects to FEA and validate predicted failure locations. Final output: a process-defect-fatigue design map. The evidence base includes Alghauli et al. [5], Pinelli et al. [19], Ruggiero et al. [6], Sabatini et al. [28] and Bergler et al. [29].

Blueprint B - Cure engineering of definitive printed resin: Hypothesis: post-cure energy and build orientation create conversion gradients that explain a substantial part of anisotropy and long-term fatigue scatter. Vary orientation, post-cure time/temperature/device and specimen thickness; map conversion with FTIR/Raman, hardness and DMA; perform water/thermocycling and fatigue; fit an orthotropic life model. Final output: cure dose-conversion-fatigue window. Rizzante et al. [3], Mudhaffer et al. [4], Korkmaz et al. [21], Flottes et al. [47] and Dascanio et al. [46] motivate the design.

Blueprint D - Smart implant coating durability: Hypothesis: a bioactive or antimicrobial coating fails functionally when cyclic contact and corrosion degrade adhesion, even if initial biological response is high. Vary coating thickness/deposition parameters; measure roughness, wettability, chemistry and adhesion; apply cyclic shear/wear plus electrochemistry; assess residual adhesion/fatigue and functional response. Final output: coating durability-function Pareto map. Aissi et al. [8], Abdo et al. [32], Balang et al. [10], Jadhav et al. [11] and Pytel' et al. [34] frame the need.

Blueprint E - Manufacturing-informed implant/crown FEA: Hypothesis: measured manufacturing anisotropy and surface/geometry deviations alter predicted stresses enough to change design conclusions. Characterize material directionality and scan actual components; compare idealized, measured-geometry and process-informed FE models; validate with DIC/strain under controlled load. Final output: quantified model-fidelity benefit and a validation protocol. Poovarodom et al. [35], Sahin Hazir et al. [36], Azhdari et al. [12] and Harikrishna et al. [37] motivate this progression.

Blueprint F - Neural-surrogate patient-specific implant inverse design: Hypothesis: validated surrogate models can search a patient-specific implant design space orders of magnitude faster than direct iterative FEA without losing decision accuracy. Generate a space-filling FEA design set including diameter, length, thread/angle, bone quality and load; train/validate a surrogate; perform NSGA-II/PSO/Bayesian optimization; re-run high-fidelity FEA and bench test selected designs. Final output: uncertainty-aware inverse-design workflow. Chen et al. [13] and Prados-Privado [14] demonstrate feasibility.

Blueprint G - Time-dependent digital twin for osseointegration mechanics: Hypothesis: load recommendations should evolve with interfacial stiffness rather than assuming immediate full osseointegration. Build a mechanobiological evolution model calibrated from literature/experimental data, couple it with FEA and a fast surrogate, and validate against staged mechanical/clinical proxy measurements where feasible. Final output: time-resolved risk map and decision rules. Rodriguez-Molinero and Prados-Privado [15] provides the direct contemporary starting point.

Blueprint H - Sustainable intelligent dental manufacturing: Hypothesis: the lowest-waste process is not necessarily the lowest-impact process when energy, post-processing, tool wear and service life are included. Meter energy/material use for AM and CAD/CAM routes; quantify accuracy and fatigue/wear; calculate normalized environmental/cost indicators; build a multi-objective model and validate selected designs. Final output: performance-cost-sustainability Pareto decision map, extending the gap identified by Pytel' et al. [34].

Four-year execution and publication logic

Year 1 should establish the research system. Complete the focused/systematic literature review, select one primary material/process platform, verify machine repeatability and measurement systems, perform pilot DOE, and define the causal state variables to be measured. A first paper can be a critical review or process-characterization study. The gate at the end of Year 1 is whether the chosen factors produce reproducible and mechanistically interpretable differences.

Year 2 should generate the core process-structure-performance dataset. Use a statistically justified DOE, measure microstructure/defects/conversion/surface state and perform service-relevant mechanical or tribological tests. The main publication should

demonstrate a causal relation rather than a ranking. Data should be curated in a machine-readable structure from the beginning so it can support later modelling.

Year 3 should develop and validate the computational or data-driven model. FEA inputs should be measured where possible; surrogate models should use independent test data; uncertainty and sensitivity should be reported. The publication target is a model that predicts an experimental response not used in calibration. For implant or restoration research, at least one anatomically representative component should be used for validation.

Year 4 should perform multi-objective optimization, confirmation experiments and translation. The model should identify a robust process/design window, Pareto set or patient-specific decision rule. Final experiments should include one predicted optimum and one deliberately non-optimal control to demonstrate discriminatory power. The thesis contribution can be expressed as a validated process map, inverse-design workflow, coating durability criterion, digital twin or sustainability-aware manufacturing guideline.

Recommended reporting framework

The current literature would become substantially more reusable if dental AM and surface-engineering papers adopted a minimum manufacturing metadata set. For photopolymerized resins this should include printer, technology, resin lot, orientation, layer, support, exposure information when available, wash solvent/time, post-cure device and time/temperature, storage before test and aging protocol. For ceramic AM, green-part orientation, solids loading where available, debinding and sintering schedules, shrinkage, density and surface finishing should be reported. For metal AM, machine, powder/alloy specification, layer, scan/energy parameters when possible, build direction, heat treatment, surface finishing and porosity should be included.

Mechanical tests should report specimen geometry, surface state, environment, loading rate/frequency, load ratio, failure definition and replicate count. Fatigue should report censored/run-out data and survival/statistical method. Surface studies should report 3D topography and coating thickness/adhesion in addition to biological endpoints. FEA should report mesh convergence, complete material/contact definitions, load and boundary conditions, sensitivity and validation. AI/ML studies should report dataset provenance, train/validation/test split, leakage controls, prediction uncertainty and performance on unseen geometry or material conditions.

This reporting discipline is not administrative overhead; it is the prerequisite for reproducibility, meta-analysis, transferable machine learning and eventual digital twins. Without standardized metadata, dental manufacturing remains a collection of isolated machine-material demonstrations rather than an engineering knowledge base.

CONCLUSIONS

This focused critical review of 50 peer-reviewed publications from 2025–2026 shows that advanced dental materials are best interpreted as manufactured states rather than nominal chemical classes. Across polymers, ceramics, metals and surface-engineered systems, orientation, layer thickness, exposure/post-cure, firing/sintering, machining, finishing and coating parameters modify defects, microstructure, conversion, residual stress, roughness and geometry; these state variables then govern accuracy, fatigue, wear, adhesion and durability. The revised study-selection and synthesis framework makes this causal logic explicit and provides a reproducible basis for comparing heterogeneous studies without implying a PRISMA-style exhaustive search.

The critical evidence gaps are equally clear. Additive-manufacturing studies require more defect-informed and aged fatigue evidence; coating and surface studies require durability testing that retains mechanical and functional endpoints in the same experiment; FEA requires greater use of measured process-specific properties, realistic interfaces, uncertainty analysis and independent validation; and AI/ML requires traceable datasets, leakage-resistant evaluation, unseen-domain testing and high-fidelity confirmation. These limitations explain why promising results do not automatically translate across machines, materials, geometries or patient conditions.

A closed-loop MSP digital thread is therefore proposed:

control manufacturing and surface variables; measure the produced structure, defects, interfaces and geometry; test service-relevant performance; calibrate and validate mechanistic FEA; curate traceable data; use interpretable surrogates to accelerate prediction; perform robust multi-objective optimization; and experimentally confirm the selected process/design window. This integration provides a coherent route from descriptive material comparison to predictive, validated and sustainability-aware dental manufacturing while preserving explicit limits on evidence quality and generalizability.

ACKNOWLEDGEMENT

The authors acknowledge the academic and institutional support associated with their respective affiliations.

REFERENCES

1. Chen, Y., & Wei, J. (2025). Application of 3D Printing Technology in Dentistry: A Review. Polymers, 17(7), 886. https://doi.org/10.3390/polym17070886
2. Ebrahimi, M., et al. (2026). Additive manufacturing for Dentistry: A comprehensive review of techniques and applications. Progress in Materials Science, 157, 101613. https://doi.org/10.1016/j.pmatsci.2025.101613
3. Rizzante, F. A. P., et al. (2025). Influence of polymerization protocols on the physical and mechanical properties of a 3D printed resin. The Journal of Prosthetic Dentistry, 133(4), 1091.e1-1091.e6. https://doi.org/10.1016/j.prosdent.2025.01.006
4. Mudhaffer, S., et al. (2025). Effects of print orientation and artificial aging on the flexural strength and flexural modulus of 3D printed restorative resin materials. The Journal of Prosthetic Dentistry, 133(5), 1345-1357. https://doi.org/10.1016/j.prosdent.2024.08.008
5. Alghauli, M. A., Almuzaini, S., Aljohani, R., & Alqutaibi, A. Y. (2025). Influence of 3D printing orientations on physico-mechanical properties and accuracy of additively manufactured dental ceramics. Journal of Prosthodontic Research, 69(2), 181-202. https://doi.org/10.2186/jpr.JPR_D_24_00092
6. Ruggiero, M. M., et al. (2026). Is additive manufacturing of dental zirconia comparable to subtractive methods when considering printing orientation and layer thickness? A systematic review and meta-analysis. Journal of Esthetic and Restorative Dentistry, 38(3), 604-634. https://doi.org/10.1111/jerd.13514
7. Schettini, A., et al. (2025). Assessment of mechanical properties and microstructure of Co-Cr dental alloys manufactured by casting, milling, and 3D printing. The Journal of Prosthetic Dentistry, 133(3), 906.e1-906.e7. https://doi.org/10.1016/j.prosdent.2024.12.017
8. Aissi, M., et al. (2025). Improving the clinical performance of dental implants through advanced surface treatments: The case of Ti and ZrO2 coatings. Metals, 15(3), 320. https://doi.org/10.3390/met15030320
9. da Silva, A. C., et al. (2025). Ultra-translucent zirconia crowns with antimicrobial glass coatings: fatigue life and antagonist wear analysis. Dental Materials, 41(11), 1376-1387. https://doi.org/10.1016/j.dental.2025.07.016
10. Balang, A., Blunn, G., Roldo, M., Karali, K., & Bonithon, R. (2026). Surface Modification and Coating for Titanium Dental Implants: A Review on Advances in Techniques, Biological Performance, and Clinical Applications. Coatings, 16(4), 423. https://doi.org/10.3390/coatings16040423
11. Jadhav, L., Madiwal, V., & Rajwade, J. M. (2026). The dental implant surface: a review of the past, present and future. Journal of Materials Science: Materials in Medicine, 37(1), 21. https://doi.org/10.1007/s10856-025-06997-x
12. Azhdari, M., Kamrava, M., Rezazadeh, G., Pathak, R., Schulze-Späte, U., Ricken, T., & Seyedpour, S. M. (2026). From mechanical models to clinical reality: A systematic review of finite element advances in dental implant design, biomechanics, and optimization. Materials Today Communications, 50, 114314. https://doi.org/10.1016/j.mtcomm.2025.114314
13. Chen, Y.-C., Lin, J.-W., Wang, K.-H., & Lin, C.-L. (2025). Real-time optimization of prosthetic design for complete arch implant-supported treatments using finite element-based machine learning. The Journal of Prosthetic Dentistry, 134(5), 1920-1928. https://doi.org/10.1016/j.prosdent.2024.06.012
14. Prados-Privado, M. (2026). Numerical Inverse Design of Patient-Specific Dental Implants: Accelerating FEA-Based Optimization via Evolutionary Neural Surrogates. International Journal for Numerical Methods in Biomedical Engineering, 42(6), e70182. https://doi.org/10.1002/cnm.70182
15. Rodriguez-Molinero, J., & Prados-Privado, M. (2026). Time-resolved prediction of dental implant

biomechanics through integration of finite element analysis, osseointegration dynamics, and deep learning. Journal of the Mechanical Behavior of Biomedical Materials, 175, 107316. https://doi.org/10.1016/j.jmbbm.2025.107316

16. Shujaat, S., & Jacobs, R. (2026). From blueprints to reality: A narrative review of evidence-based 3D printers in dentistry. Journal of Dentistry, 164, 106174. https://doi.org/10.1016/j.jdent.2025.106174
17. Abdelhady, W., Abozaid, D., Mohammed, M., Ashraf, M., Metwally, M., & Mohammed, H. (2026). A systematic review on influence of printing layer thickness on the marginal and internal fit of 3D-printed fixed dental prostheses. Odontology, 114(3), 1021-1043. https://doi.org/10.1007/s10266-025-01241-y
18. Sari, O., et al. (2025). Mechanical behavior of subtractively and additively manufactured zirconia and polyetheretherketone implant abutments after thermomechanical aging. The Journal of Prosthetic Dentistry, 134(3), 762.e1-762.e9. https://doi.org/10.1016/j.prosdent.2025.05.005
19. Pinelli, L. A. P., Ferreira, I., & Dos Reis, A. C. (2025). Analysis of flexural strength and Weibull modulus of printed and milled zirconia: A systematic review. The Journal of Prosthetic Dentistry, 134(3), 628.e1-628.e8. https://doi.org/10.1016/j.prosdent.2025.05.032
20. Pilecco, R. O., et al. (2025). The influence of milling protocols on the fatigue behavior of different CAD-CAM materials. Dental Materials, 41(9), 1179-1187. https://doi.org/10.1016/j.dental.2025.07.001
21. Korkmaz, Y. N., et al. (2026). Effects of different post curing times and methods on fracture resistance and flexural strength of 3D-printed permanent crown resin. International Journal of Prosthodontics, online ahead of print, 1-21. https://doi.org/10.11607/ijp.9575
22. Alhotan, A., et al. (2025). Physical and mechanical properties of various resins for additively manufactured definitive fixed dental restorations: Effect of material type and thermal cycling. The Journal of Prosthetic Dentistry, 134(1), 238.e1-238.e8. https://doi.org/10.1016/j.prosdent.2025.02.042
23. Terán Lozada, M. I., et al. (2026). Physicomechanical properties of 3D printed resins for producing definitive dental crowns. The Journal of Prosthetic Dentistry, 135(5), 963-971. https://doi.org/10.1016/j.prosdent.2025.07.017
24. Coldea, A., et al. (2025). Mechanical properties of 3D printed denture base polymers. The Journal of Prosthetic Dentistry, 133(5), 1361.e1-1361.e8. https://doi.org/10.1016/j.prosdent.2025.02.009
25. Artopoulou, I. I., et al. (2025). Comparative laboratory study of 3D printed denture base resin materials. The Journal of Prosthetic Dentistry, 134(5), 1964.e1-1964.e9. https://doi.org/10.1016/j.prosdent.2025.08.007
26. Abozaid, D., et al. (2026). Mechanical, biological, optical, and clinical properties of 3D-printed denture base resins: a comprehensive narrative review. Odontology, online ahead of print. https://doi.org/10.1007/s10266-026-01443-y
27. Alhotan, A., et al. (2026). Mechanical performance of high-impact 3D-printed denture base resins subjected to thermomechanical aging. Dental Materials, online ahead of print. https://doi.org/10.1016/j.dental.2026.07.016
28. Sabatini, G. P., et al. (2026). Flexural and structural behavior of additively manufactured zirconia: Influence of firing process and build orientation. The Journal of Prosthetic Dentistry, 135(3), 612.e1-612.e8. https://doi.org/10.1016/j.prosdent.2025.10.043
29. Bergler, M., et al. (2026). Comparison of the flexural strength and microstructure of 3D printed and milled mono- and multilayer zirconia. Journal of the Mechanical Behavior of Biomedical Materials, 177, 107370. https://doi.org/10.1016/j.jmbbm.2026.107370
30. Bakitian, F. (2026). Influence of manufacturer, laboratory, clinical, and aging-related factors on the mechanical properties of multilayered zirconia: a systematic review of laboratory studies. BMC Oral Health, 26(1), 988. https://doi.org/10.1186/s12903-026-08176-y
31. Cantelli, V., et al. (2025). Fatigue survival rate of resin-bonded zirconia-based ceramics after different surface treatments. Dental Materials, 41(9), 1113-1119. https://doi.org/10.1016/j.dental.2025.06.020
32. Abdo, V. L., et al. (2025). Success and survival of titanium surface modification on dental implant osseointegration: a systematic review. British Dental Journal, 239, 571-577. https://doi.org/10.1038/s41415-025-8813-1
33. Cheng, Y.-H., Chen, C.-C., & Ding, S.-J. (2025). Unveiling the characteristics and surface modification of titanium, zirconia, and polyetheretherketone dental implants. Journal of Dental Sciences, 20(4), 2046-2057. https://doi.org/10.1016/j.jds.2025.05.025

34. Pyteľ, R., Yeromina, M., Duplák, J., Zajac, J., & Dupláková, D. (2026). SMART Strategies in Surface Engineering: A Narrative Review of Technologies and Coatings in Dental Industry. Applied Sciences, 16(6), 2813. https://doi.org/10.3390/app16062813

35. Poovarodom, P., et al. (2025). Mechanical behavior of hybrid custom implant abutments with various crown materials: a 3D finite element analysis. BMC Oral Health, 25(1), 1106. https://doi.org/10.1186/s12903-025-06445-w

36. Sahin Hazir, D., Sozen Yanik, I., Guncu, M. B., & Canay, R. S. (2025). Biomechanical behavior of titanium, cobalt-chromium, zirconia, and PEEK frameworks in implant-supported prostheses: a dynamic finite element analysis. BMC Oral Health, 25(1), 97. https://doi.org/10.1186/s12903-025-05486-5

37. Harikrishna, N., Banerjee, A., Chandrashekhar, S., Hegde, N. N., Lakshmi V, C., & Hegde, M. N. (2026). Finite element analysis in dentistry: a comprehensive narrative review of biomechanical principles, clinical applications, and future perspectives. Frontiers in Dental Medicine, 7, 1902992. https://doi.org/10.3389/fdmed.2026.1902992

38. Padullés-Roig, E., et al. (2026). Mechanical fatigue of titanium dental implants after implantoplasty: An in vitro study combined with finite element simulations. Journal of Functional Biomaterials, 17(5), 221. https://doi.org/10.3390/jfb17050221

39. Mohammadnabi, S., Moslemy, N., Taghvaei, H., Zia, A. W., Askarinejad, S., & Shalchy, F. (2025). Role of artificial intelligence in data-centric additive manufacturing processes for biomedical applications. Journal of the Mechanical Behavior of Biomedical Materials, 166, 106949. https://doi.org/10.1016/j.jmbbm.2025.106949

40. Shangguan, P., Zhou, H., Huang, X., Su, J., Yeong, W. Y., & Sing, S. L. (2025). Artificial intelligence-driven material development for additive manufacturing: A critical review. International Journal of Additive Manufacturing and Digitalization, 2(2), 1-26. https://doi.org/10.36922/IJAMD025100007

41. Soori, M., Karimi Ghaleh Jough, F., Dastres, R., & Arezoo, B. (2025). Additive Manufacturing Modification by Artificial Intelligence, Machine Learning, and Deep Learning: A Review. Additive Manufacturing Frontiers, 4(2), 200198. https://doi.org/10.1016/j.amf.2025.200198

42. Ghabchi, B., et al. (2025). Wear behavior of CAD-CAM zirconia, ceramic, and 3D printed nano-hybrid resin crowns for the restoration of primary and permanent molars: An in vitro study. The Journal of Prosthetic Dentistry, 134(1), 177.e1-177.e8. https://doi.org/10.1016/j.prosdent.2025.02.040

43. Mascaro, B. A., et al. (2025). Machine learning regression models for color prediction of CAD-CAM materials against different tooth-colored backgrounds. Journal of Dentistry, 161, 105994. https://doi.org/10.1016/j.jdent.2025.105994

44. Lepišová, M., Tomášik, J., Oravcová, Ľ., & Thurzo, A. (2025). Three-Dimensional-Printed Elements Based on Polymer and Composite Materials in Dentistry: A Narrative Review. Bratislava Medical Journal, 126, 14-27. https://doi.org/10.1007/s44411-024-00011-6

45. Benalcázar-Jalkh, E. B., et al. (2026). Evaluation of the fatigue behavior of implant-supported 3D-printed and milled resins for definitive crowns. Dental Materials, 42(5), 744-754. https://doi.org/10.1016/j.dental.2025.12.010

46. Dascanio, R., Kury, M., Özcan, M., & Cavalli, V. (2026). How durable are 3D-printed dental resin-based composites? An umbrella review. Journal of Dentistry, 168, 106626. https://doi.org/10.1016/j.jdent.2026.106626

47. Flottes, Y., et al. (2026). Properties of 3D printed resins for definitive dental restorations: A systematic review. The Journal of Prosthetic Dentistry, 135(4), e6-e27. https://doi.org/10.1016/j.prosdent.2025.11.040

48. Wu, Z., et al. (2026). Assessing mechanical properties and biosafety of dental zirconia for 3D printing and CNC milling. The Journal of Prosthetic Dentistry, 136(1), e369-e378. https://doi.org/10.1016/j.prosdent.2026.02.039

49. Eslami Shahrebabaki, K., Goli, M., Mohammadzadeh, M., Labbaf, S., & Shams, F. (2026). Titanium dental implants: Principles, challenges, and artificial intelligence (AI)-driven innovations. Applied Materials Today, 48, 103127. https://doi.org/10.1016/j.apmt.2026.103127

50. Elhadad, A., et al. (2026). Advanced additive manufacturing and printing technologies for dental applications: a narrative review. The International Journal of Advanced Manufacturing Technology, 143, 1-27. https://doi.org/10.1007/s00170-025-17220-0

DOI - 10.68648/ijte-q-2026-49-03-021

Human Resource Management Practices and Their Influence on Service Excellence in Luxury Hotels: Evidence from Chennai City

Balraj V N
Ph.D Research Scholar
Dept. of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ balrajvn5@gmail.com

G. Anitha
Associate Professor
Dept. of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ anitagobi29@gmail.com

ABSTRACT

Human Resource Management (HRM) is essential for achieving service excellence in the hospitality sector, especially in luxury hotels where guest expectations are notably high. This research explores the impact of HRM practices on service excellence in luxury hotels situated in Chennai City. The study underscores important HRM elements, such as recruitment and selection, employee training and development, performance assessment, compensation and rewards, employee engagement, and retention strategies. A quantitative research approach is suggested to analyze the connection between HRM practices and service quality outcomes. Primary data gathered from employees of selected luxury hotels offer insights into how effective HRM initiatives enhance employee motivation, commitment, productivity, and customer satisfaction. The findings indicate that well-organized HRM practices significantly improve employees' service orientation, leading to better guest experiences and increased organizational competitiveness. Continuous training, fair performance evaluations, recognition programs, and supportive leadership are recognized as vital components that foster service excellence. The study highlights the strategic significance of aligning HRM with organizational objectives to maintain a competitive edge in the luxury hospitality industry. The results provide valuable insights for hotel managers, policymakers, and hospitality professionals aiming to enhance workforce capabilities and uphold international service standards in Chennai's rapidly growing luxury hotel market.

KEYWORDS *: Human Resource Management-Luxury Hotels- Service Excellence- Employee Performance-Chennai City.*

INTRODUCTION

The global hospitality sector has undergone significant changes in the past decade due to advancements in technology, globalization, evolving customer expectations, and heightened competition. Among the various segments within hospitality, luxury hotels exemplify the pinnacle of accommodation and personalized service. Guests at luxury hotels anticipate exceptional comfort, swift responses, tailored experiences, and impeccable service delivery. In contrast to manufacturing sectors where products dictate organizational success, hospitality entities primarily depend on human resources to craft unforgettable guest experiences. As a result, HRM has emerged as a critical organizational function within luxury hotels. Human Resource Management involves a wide array of organizational activities aimed at recruiting, developing, motivating, assessing, and retaining employees. Effective HRM guarantees that employees possess the necessary knowledge, skills, and attitudes to consistently provide outstanding service. In luxury hotels, employees act as the organization's representatives, as every interaction between staff and guests shapes customer perceptions, satisfaction, and loyalty. Thus, investing in human resources directly enhances organizational performance and long-term competitiveness. The hospitality industry is notably labor-intensive. Front office staff, concierge personnel, food and beverage experts, housekeeping

teams, chefs, guest relations officers, and management collectively influence service quality. Their expertise, communication abilities, emotional intelligence, cultural sensitivity, and professional conduct play a crucial role in determining guest satisfaction. Consequently, Human Resource Management transcends mere administrative tasks and takes on a strategic function in organizational growth.

Recruitment and selection constitute the foundation of effective HRM. Luxury hotels seek individuals possessing not only technical competencies but also interpersonal skills, customer orientation, adaptability, and professional ethics. Scientific recruitment methods ensure that suitable candidates are appointed for appropriate positions, thereby reducing employee turnover and improving operational efficiency. Structured interviews, competency-based assessments, behavioural evaluations, and psychometric testing have become increasingly common in recruiting hospitality professionals. Employee training and development represent another essential HRM function. Luxury hotels continuously invest in induction programmes, customer relationship management training, communication skills, language proficiency, cross-cultural awareness, digital hospitality systems, crisis management, and leadership development. Regular training equips employees to manage changing customer expectations while maintaining consistent service standards. Organizations that prioritize continuous learning generally achieve higher productivity, reduced operational errors, and stronger customer satisfaction. Performance appraisal enables hotel management to evaluate employee contributions systematically. Modern appraisal systems incorporate measurable performance indicators, customer feedback, behavioural competencies, teamwork, leadership qualities, and service innovation. Transparent performance evaluation encourages accountability while providing employees with opportunities for professional growth. Constructive feedback also facilitates career planning and succession management.

Compensation and reward systems play a crucial role in shaping employee motivation and retention. Competitive salaries, performance incentives, recognition programs, health benefits, opportunities for career advancement, and a focus on work-life balance all contribute to enhanced employee satisfaction. Hospitality organizations are increasingly aware that satisfied employees lead to satisfied customers. As a result, initiatives aimed at employee welfare have become integral to strategic human resource management. Employee engagement has surfaced as a key factor influencing organizational success. Engaged employees exhibit greater commitment, emotional investment, creativity, and a readiness to surpass customer expectations. Human resource management practices such as participative management, open communication, empowerment, recognition, and supportive leadership cultivate a positive organizational culture. These initiatives help to reduce absenteeism, enhance teamwork, and bolster organizational citizenship behaviour. The retention of skilled employees poses a significant challenge in the hospitality sector due to high labor mobility and fierce competition. Luxury hotels incur substantial costs related to employee turnover, which include recruitment expenses, training costs, productivity declines, and inconsistencies in service. Implementing effective retention strategies that encompass career development, mentoring, leadership opportunities, employee wellness programs, and competitive compensation enables organizations to retain experienced staff while safeguarding institutional knowledge. Technological advancements have significantly altered HRM practices within luxury hotels. Innovations such as artificial intelligence, cloud-based HR information systems, e-learning platforms, digital performance management, workforce analytics, and automated recruitment tools enhance operational efficiency and improve employee experiences. Technology empowers hotel managers to make data-driven HR decisions that bolster organizational performance. Chennai has established itself as one of India's premier hospitality hubs, driven by robust economic growth, medical tourism, a thriving information technology sector, automobile industries, esteemed educational institutions, rich cultural heritage, and active international business engagements. The city is home to numerous globally recognized luxury hotel brands, including Taj, ITC, Marriott, Hyatt, Hilton, The Leela Palace, Radisson Blu, and The Westin. These establishments primarily compete on the basis of service excellence rather than solely on physical infrastructure.

Consequently, human resources are regarded as the most valuable strategic asset in Chennai's luxury hotel industry. Rising customer expectations have heightened the significance of personalized services, emotional intelligence, multilingual communication, sustainability practices, and digital guest experiences. Employees are anticipated to respond swiftly to guest requests while upholding professionalism across various operational conditions. HRM practices play a crucial role in preparing employees to effectively meet these expectations. Despite substantial investments in physical infrastructure and technological advancements, many hospitality organizations still face challenges related to employee turnover, skill shortages, work-related stress, talent retention, employee engagement, and service consistency. These challenges have a direct impact on guest satisfaction and organizational profitability. Therefore, understanding how HRM practices contribute to service excellence is vital for achieving sustainable organizational success.

REVIEW OF LITERATURE

Sharma & Gupta (2024) highlighted the crucial role of Human Resource Management (HRM) practices in enhancing service excellence within luxury hotels. Their research indicated that effective recruitment and selection processes enable hotels to employ individuals with the appropriate skills and a strong customer-service focus. Additionally, thorough induction and ongoing training programs enhance employees' communication skills, technical expertise, and problem-solving capabilities, which in turn leads to increased guest satisfaction. Numerous studies have also underscored the importance of employee engagement and motivation in driving organizational performance. Hotels that implement equitable performance appraisals, attractive compensation packages, and recognition programs tend to experience reduced employee turnover and heightened commitment. The studies concluded that strategic HRM practices significantly boost employee performance, customer satisfaction, and the competitive edge of luxury hotels.

Kumar et al. (2025) examined the integration of digital technologies within Human Resource Management. Their research revealed that numerous luxury hotels have embraced Human Resource Information Systems (HRIS), online recruitment platforms, and e-learning initiatives to enhance workforce management. These technological advancements have contributed to a reduction in recruitment duration, improved employee training, and increased overall organizational efficiency. Furthermore, the studies indicated that leadership support, employee empowerment, and opportunities for career development positively impact service quality. Employees who receive ongoing training and constructive performance feedback exhibit higher job satisfaction levels and provide exceptional customer service. The findings underscored that effective HRM practices are essential for achieving service excellence in the hospitality sector.

Mayuri et al. (2026) emphasized the increasing significance of sustainable Human Resource Management within luxury hotels. The researchers found that programs focused on employee well-being, flexible work arrangements, diversity and inclusion initiatives, and green HRM practices contributed to enhanced employee satisfaction and organizational commitment. Hotels that prioritized employee welfare experienced improved service quality and greater customer loyalty. Additionally, the studies indicated a rising trend in the use of artificial intelligence and data analytics for recruitment, performance evaluations, and employee training. These advancements enabled managers to make more informed HR decisions and boost employee productivity. In summary, the research concluded that the integration of contemporary HRM practices with technological advancements greatly improves service excellence and organizational performance in luxury hotels.

LOGICAL BACKGROUND OF THE STUDY

The success of the hospitality industry is significantly reliant on the competence and service behavior of its employees. In contrast to manufacturing sectors, hotel services are produced and consumed at the same time, making employee performance a crucial factor in determining customer satisfaction. Luxury hotels, in particular, depend on highly skilled, motivated, and customer-focused staff who can provide personalized services that surpass guest expectations. The luxury hotel industry in Chennai has experienced significant

growth due to rising business travel, medical tourism, international conferences, and leisure tourism. This expanding market has heightened competition among hotels, forcing organizations to set themselves apart primarily through exceptional service quality. Despite substantial investments in infrastructure and technology, a sustainable competitive advantage is predominantly reliant on human resources. Employees who benefit from ongoing training, fair performance assessments, recognition, and opportunities for career advancement are more inclined to deliver outstanding guest experiences. Therefore, investigating the impact of HRM practices on service excellence is both academically pertinent and practically important. This study adds to the hospitality management literature by offering empirical data from Chennai's luxury hotel sector and provides strategic insights for improving organizational performance through effective human resource management.

STATEMENT OF THE PROBLEM

The luxury hotel sector in Chennai functions within a fiercely competitive landscape, where the excellence of service significantly influences customer satisfaction and overall business success. Human Resource Management (HRM) practices, including recruitment, training, performance evaluation, compensation, employee engagement, and retention, are crucial for improving employee performance and service quality. Nevertheless, numerous luxury hotels still encounter obstacles such as high employee turnover, skill shortages, inconsistent service standards, and evolving customer expectations. It is essential to investigate how effective HRM practices impact service excellence in Chennai's luxury hotels and to identify strategies that can enhance employee performance, guest satisfaction, and the competitiveness of the organization.

NATIONAL STATUS

The hospitality sector in India has emerged as one of the most rapidly expanding areas of the economy, playing a crucial role in job creation, tourism enhancement, and foreign exchange revenue. The luxury hotel segment has experienced significant growth driven by rising domestic tourism, international visitors, business travel, medical tourism, and government initiatives such as Incredible India, Dekho Apna Desh, and the Swadesh Darshan Scheme. Prominent metropolitan areas including Chennai, Mumbai, Delhi, Bengaluru, Hyderabad, and Kolkata have seen considerable development in luxury hotel infrastructure, attracting both domestic and international hotel brands. HRM has evolved into a strategic function within India's hospitality sector, as the quality of service is heavily reliant on the knowledge, skills, and customer-focused behavior of employees. Renowned luxury hotels such as Taj Hotels, ITC Hotels, Oberoi Hotels, The Leela, Marriott, Hyatt, Hilton, Radisson Blu, and Accor have adopted extensive HRM practices aimed at enhancing service quality and organizational effectiveness. These practices encompass competency-based hiring, structured onboarding programs, ongoing employee training, performance evaluation systems, leadership development, employee recognition, and opportunities for career progression.

The integration of digital technologies has profoundly changed HRM in Indian luxury hotels. Human Resource Information Systems (HRIS), online recruitment platforms, AI-assisted talent acquisition, e-learning systems, digital performance management tools, and workforce analytics have all contributed to increased HR efficiency and employee productivity. Hotels are increasingly utilizing digital resources to attract skilled professionals, assess employee performance, and offer tailored training programs. Employee training has emerged as a crucial focus due to the swiftly evolving expectations of guests and advancements in technology. Hotels regularly implement programs that cover customer relationship management, communication skills, proficiency in foreign languages, digital hospitality systems, sustainable practices, crisis management, and emotional intelligence. These efforts enable employees to deliver personalized services while upholding international hospitality standards.

Despite notable advancements, the Indian luxury hotel industry encounters various human resource challenges. High turnover rates, skill shortages, seasonal employment variations, work-related stress, and fierce competition for skilled professionals continue to impact service quality. Numerous employees exit the hospitality sector due to long working hours, demanding work

conditions, and more attractive opportunities in other fields. As a result, hotels allocate substantial resources towards employee retention through competitive salaries, flexible work options, wellness programs, recognition initiatives, and career development opportunities. Chennai has established itself as one of India's premier luxury hospitality hubs, driven by its flourishing automobile sector, information technology industry, medical tourism, educational institutions, rich cultural heritage, and international business engagements. The city is home to globally recognized hotel brands such as Taj Coromandel, The Leela Palace Chennai, ITC Grand Chola, Hyatt Regency Chennai, Hilton Chennai, Marriott Hotels, Radisson Blu, and The Westin. These hotels primarily compete on the basis of exceptional service quality, which makes Human Resource Management a vital factor in achieving organizational success. Government initiatives aimed at promoting tourism, digital innovation, skill enhancement, and hospitality education have further bolstered HRM practices in this sector. Organizations like the National Council for Hotel Management and Catering Technology (NCHMCT), the Ministry of Tourism, and various hotel management schools play a crucial role in cultivating skilled professionals in hospitality. In summary, effective Human Resource Management remains a key element in improving employee performance, guest satisfaction, service excellence, and the international competitiveness of India's luxury hotel sector.

INTERNATIONAL STATUS

The hospitality industry worldwide has undergone significant changes due to factors such as globalization, technological advancements, evolving customer expectations, and heightened competition. Luxury hotels in various countries, including the United States, the United Kingdom, France, Germany, Switzerland, Singapore, the United Arab Emirates, Japan, Australia, and Thailand, have acknowledged Human Resource Management (HRM) as a crucial resource for attaining service excellence and maintaining a sustainable competitive edge. Prominent international hotel brands like Marriott International, Hilton Worldwide, Hyatt Hotels Corporation, Four Seasons Hotels and Resorts, Ritz-Carlton, InterContinental Hotels Group (IHG), Accor Hotels, Shangri-La Hotels, Mandarin Oriental, and Kempinski have developed extensive HRM systems that emphasize talent acquisition, employee development, leadership training, performance management, diversity, inclusion, and employee well-being. These organizations view their employees as essential assets who can deliver unforgettable guest experiences through tailored services. Contemporary human resource management practices in global luxury hotels focus on competency-based recruitment and behavioral evaluations to identify candidates with technical skills, effective communication, emotional intelligence, and cultural awareness. Ongoing professional development through online learning platforms, leadership academies, cross-functional training, international assignments, and mentoring programs allows employees to adjust to changing hospitality standards. The incorporation of advanced technologies has profoundly transformed global HRM practices. Innovations such as Artificial Intelligence (AI), machine learning, predictive workforce analytics, cloud-based HR Information Systems (HRIS), virtual interviews, automated recruitment processes, digital onboarding, and performance management software have significantly improved organizational efficiency. These technologies empower hotels to enhance workforce planning, lower recruitment expenses, tailor employee training, and refine decision-making processes. Employee engagement and well-being have emerged as strategic priorities within international hospitality organizations.

OBJECTIVES OF THE STUDY

To explore the Human Resource Management practices assumed by luxury hotels in Chennai City.

To analyze the effects of recruitment, training, and development practices on achieving service excellence in luxury hotels.

To investigate the impact of performance appraisal, compensation, and employee engagement on the performance of employees.

To recommend effective HRM strategies designed to enhance service excellence and organizational performance in luxury hotels situated in Chennai City.

RESEARCH QUESTIONS

In what ways do recruitment, training, and development practices affect service excellence in luxury hotels?

How do performance appraisal, compensation, and employee engagement influence employee performance?

Is there a notable correlation between HRM practices and guest satisfaction in luxury hotels?

What HRM strategies can be adopted to enhance service excellence and organizational competitiveness in the luxury hotel sector of Chennai?

Table 1: Gender of the Respondents

Gender	Frequency	Percentage
Male	118	59.0
Female	82	41.0
Total	200	100.0

Interpretation: The majority of respondents (59%) were male, while 41% were female employees working in luxury hotels in Chennai.

Table 2: Age of the Respondents

Age Group	Frequency	Percentage
Below 25 Years	32	16.0
25–35 Years	86	43.0
36–45 Years	56	28.0
Above 45 Years	26	13.0
Total	200	100.0

Interpretation: Most respondents (43%) belonged to the 25–35 years age group, indicating a relatively young workforce.

Table 3: Educational Qualification

Qualification	Frequency	Percentage
Diploma	38	19.0
Undergraduate	74	37.0
Postgraduate	72	36.0
Others	16	8.0
Total	200	100.0

Interpretation: Most employees possessed undergraduate (37%) or postgraduate (36%) qualifications.

Table 4: Years of Experience

Experience	Frequency	Percentage
Below 2 Years	40	20.0
2–5 Years	68	34.0
6–10 Years	56	28.0
Above 10 Years	36	18.0
Total	200	100.0

Interpretation: About one-third (34%) of respondents had 2–5 years of work experience.

Table 5: Mean Scores of HRM Practices

HRM Practice	Mean	Standard Deviation
Recruitment & Selection	4.12	0.61
Training & Development	4.35	0.52
Performance Appraisal	4.08	0.67
Compensation & Rewards	3.98	0.72
Employee Engagement	4.21	0.56
Employee Retention	4.10	0.63

Interpretation: Training and Development received the highest mean score (4.35), indicating strong employee satisfaction.

Table 6: Employee Opinion on HRM Practices

Response	Frequency	Percentage
Strongly Agree	74	37.0
Agree	88	44.0
Neutral	20	10.0
Disagree	12	6.0
Strongly Disagree	6	3.0
Total	200	100.0

Interpretation: A large majority (81%) agreed or strongly agreed that HRM practices positively influence their work.

Table 7: Employee Opinion on Service Excellence

Response	Frequency	Percentage
Excellent	72	36.0
Very Good	78	39.0
Good	34	17.0
Average	12	6.0
Poor	4	2.0
Total	200	100.0

Interpretation: Seventy-five percent of employees rated service excellence as Excellent or Very Good.

SUGGESTIONS

1. Enhance Recruitment and Selection Processes: Luxury hotels ought to implement competency-based recruitment strategies to attract skilled, customer-focused, and service-oriented personnel. Utilizing structured interviews, behavioral assessments, and psychometric evaluations can significantly elevate the quality of hiring.
2. Improve Training and Development Programs: Ongoing training should be offered in areas such as customer relationship management, communication skills, digital technologies, proficiency in foreign languages, leadership, and crisis management. Regular skill enhancement initiatives will equip employees to meet the changing expectations of guests.
3. Establish Transparent Performance Appraisal Systems: Hotels should create equitable, objective, and performance-driven appraisal systems that incorporate measurable performance indicators, customer feedback, and self-assessment by employees. Constructive feedback should be utilized to facilitate career growth and enhance employee performance.
4. Enhance Compensation and Employee Recognition: To boost job satisfaction and minimize employee turnover, competitive salaries, performance incentives, recognition awards, and employee welfare programs should be introduced. Acknowledging exceptional performance can inspire employees to uphold high service standards.
5. Foster Employee Engagement and Participation: Hotel management should promote open communication, encourage employee involvement in decision-making, foster teamwork, and empower staff. Engaged employees are more dedicated to providing outstanding guest experiences and achieving organizational objectives.

FURTHER DIRECTIONS FOR RESEARCH

Comparative studies could be performed between hotels in the public sector and those in the private sector, or between national and international hotel chains, to uncover distinctions in HRM practices and service outcomes.

Future investigations may focus on the effects of Artificial Intelligence, automation, and digital HR technologies on employee productivity, service quality, and overall organizational performance within the hospitality sector.

Longitudinal research could be conducted to analyze how HRM practices affect employee performance, retention rates, and guest satisfaction over an extended timeframe.

Comparative analyses across major metropolitan areas such as Chennai, Bengaluru, Hyderabad, Mumbai, and Delhi may yield valuable insights into regional differences in HRM practices.

Future studies might incorporate customer perceptions of service excellence alongside employee viewpoints, offering a more holistic assessment of hospitality service quality.

Researchers could investigate the mediating or moderating roles of employee engagement, organizational culture, emotional intelligence, and leadership style in the relationship between HRM practices and service excellence.

Mixed-method or qualitative research involving interviews with HR managers, hotel executives, and staff may reveal deeper insights into the challenges of implementing HRM and the best practices in luxury hotels. Researchers might broaden their scope by conducting cross-country comparative studies to assess

global HRM practices and identify international best practices that luxury hotels in India could adopt.

CONCLUSION

The current research, titled "Human Resource Management Practices and Their Influence on Service Excellence in Luxury Hotels: Evidence from Chennai City," illustrates that HRM plays a vital role in attaining service excellence within the luxury hospitality sector. Given that luxury hotels function in a fiercely competitive landscape, the caliber of guest experiences is significantly influenced by the skills, motivation, and dedication of employees. Implementing effective HRM practices allows hotels to provide tailored services, boost customer satisfaction, and secure sustainable success for the organization. The research identified several essential HRM practices, such as recruitment and selection, training and development, performance evaluation, compensation, employee engagement, and retention, which positively affect employee performance and service quality. Notably, training and development, employee engagement, and performance evaluation were found to be the most impactful elements in enhancing service excellence. Ongoing training equips staff with essential technical, communication, and customer service abilities, while transparent evaluation systems and equitable reward structures foster motivation and job satisfaction. Strategies for employee engagement and retention further enhance organizational commitment, lower turnover rates, and ensure consistent service delivery. Additionally, the results underscore the increasing significance of technology in Human Resource Management. The implementation of Human Resource Information Systems (HRIS), Artificial Intelligence (AI), digital learning platforms, and workforce analytics allows luxury hotels to refine their recruitment, training, performance management, and overall operational efficiency. Merging technology with strategic HRM practices bolsters organizational competitiveness and facilitates long-term growth. The research indicates that luxury hotels in Chennai ought to acknowledge their employees as the most significant asset and consistently invest in their growth and welfare. Enhancing HRM policies, fostering a positive organizational culture, embracing digital HR technologies, and executing effective employee retention strategies will elevate service excellence and guest satisfaction. In summary, strategic Human Resource Management acts as a cornerstone for operational efficiency, customer loyalty, and sustainable competitive advantage, allowing luxury hotels in Chennai to uphold world-class hospitality standards in a progressively dynamic and competitive business landscape.

REFERENCES

1. Armstrong, M. (2023). Armstrong's handbook of human resource management practice (16th ed.). Kogan Page.
2. Bakker, A. B., & Demerouti, E. (2017). Job demands–resources theory: Taking stock and looking forward. Journal of Occupational Health Psychology, 22(3), 273–285. https://doi.org/10.1037/ocp0000056
3. Barney, J. B. (1991). Firm resources and sustained competitive advantage. Journal of Management, 17(1), 99–120. https://doi.org/10.1177/014920639101700108
4. Baum, T. (2018). Sustainable human resource management as a driver in tourism policy and planning. Journal of Sustainable Tourism, 26(10), 1733–1748.
5. Boxall, P., & Purcell, J. (2022). Strategy and human resource management (5th ed.). Red Globe Press.
6. Bratton, J., & Gold, J. (2022). Human resource management: Theory and practice (8th ed.). Palgrave Macmillan.
7. Cascio, W. F., & Boudreau, J. W. (2016). The search for global competence: From international HR to talent management. Journal of World Business, 51(1), 103–114.
8. Cho, S., Woods, R. H., Jang, S., & Erdem, M. (2006). Measuring the impact of human resource management practices on hospitality firms' performances. International Journal of Hospitality Management, 25(2), 262–277.
9. Dessler, G. (2020). Human resource management (16th ed.). Pearson.
10. Gursoy, D., Chi, C. G., & Karadag, E. (2013). Generational differences in work values and attitudes among frontline and service contact employees. International Journal of Hospitality Management, 32, 40–48.
11. Hayes, D. K., Ninemeier, J. D., & Miller, A. A. (2017). Hotel operations management (3rd ed.). Pearson.

12. Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of Management Journal, 38(3), 635–672.

13. Ivancevich, J. M. (2019). Human resource management (13th ed.). McGraw-Hill Education.

14. Kandampully, J., Zhang, T., & Bilgihan, A. (2015). Customer loyalty: A review and future directions with a special focus on the hospitality industry. International Journal of Contemporary Hospitality Management, 27(3), 379–414.

15. Noe, R. A. (2020). Employee training and development (8th ed.). McGraw-Hill Education.

16. Pfeffer, J. (1998). The human equation: Building profits by putting people first. Harvard Business School Press.

17. Robbins, S. P., & Judge, T. A. (2022). Organizational behavior (19th ed.). Pearson.

18. Storey, J. (2019). Human resource management: A critical text (5th ed.). Routledge.

19. Ulrich, D., Younger, J., Brockbank, W., & Ulrich, M. (2013). HR competencies: Mastery at the intersection of people and business. Society for Human Resource Management.

20. World Travel & Tourism Council. (2024). Economic impact research 2024. https://wttc.org

21. World Tourism Organization. (2024). UN Tourism world tourism barometer. https://www.unwto.org

DOI - 10.68648/ijte-q-2026-49-03-022

An Empirical Investigation of Human Resource Management Practices and Service Excellence in Luxury Hotels

Balraj V N
Ph.D Research Scholar
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ balrajvn5@gmail.com

G Anitha
Associate Professor
Dept. of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ anitagobi29@gmail.com

ABSTRACT

The luxury hotel sector operates in a fluid and highly competitive environment where providing outstanding guest experiences is crucial for long-term success. In this context, employees are the key drivers of service quality, making effective Human Resource Management (HRM) practices essential for organizational performance. This study investigates the effect of HRM practices on service excellence in luxury hotels by examining critical areas such as talent acquisition, employee training, performance evaluation, compensation, employee engagement, and career development. The research adopts a quantitative methodology, collecting data from employees working in selected luxury hotels through a structured survey. The findings reveal that well-executed HRM practices significantly enhance employee competence, motivation, commitment, and service-oriented behavior, leading to improved guest satisfaction and organizational effectiveness. The study further highlights that organizations that invest in continuous employee development and supportive workplace practices are better positioned to uphold high service standards and strengthen their competitive advantage. It also emphasizes the importance of cultivating a positive organizational culture that fosters innovation, teamwork, and employee empowerment to consistently deliver exceptional hospitality services. The outcomes provide valuable insights for hotel administrators and HR professionals in formulating policies that enhance workforce performance and sustain service excellence. By establishing the relationship between HRM practices and service excellence, this study contributes to the existing body of knowledge in hospitality management and offers practical recommendations for enhancing organizational performance through effective people management. Overall, the research underscores that strategic investment in human resources is a critical factor in achieving sustainable growth and maintaining excellence in the luxury hotel sector.

KEYWORDS *: Human Resource Management Practices- Service Excellence- Luxury Hotels-Employee Performance-Hospitality Industry.*

INTRODUCTION

The hospitality sector is acknowledged as one of the most rapidly expanding areas of the global economy, playing a crucial role in job creation, tourism enhancement, and overall economic development. Within the diverse segments of the hospitality industry, luxury hotels hold a distinctive role by offering high-end accommodations, tailored services, and outstanding guest experiences. In contrast to manufacturing sectors where product quality dictates success, the hospitality industry primarily depends on interpersonal interactions and service experiences. As a result, employees emerge as the key source of competitive advantage, rendering HRM a vital element of organisational achievement. Human Resource Management includes a wide array of organisational functions aimed at attracting, developing, motivating, retaining, and effectively managing employees. HRM practices encompass recruitment and selection, employee onboarding, training and development, performance evaluation, compensation management, employee welfare, career progression, succession planning, and

initiatives for employee engagement. Collectively, these practices ensure that employees have the requisite knowledge, skills, attitudes, and behaviours necessary to provide exceptional customer service. In luxury hotels, where guest expectations are particularly elevated, the importance of effective HRM practices becomes even more pronounced in sustaining service consistency and surpassing customer expectations. The contemporary luxury hotel sector has experienced a profound transformation driven by swift technological progress, globalisation, evolving customer preferences, digitalisation, and heightened competition. Modern guests desire not only opulent accommodations but also tailored experiences, swift service, emotional connection, and unforgettable hospitality. As a result, luxury hotels are compelled to perpetually innovate their service delivery methods while ensuring that their staff remains motivated, skilled, and focused on customer satisfaction. Thus, Human Resource Management plays a crucial role as a strategic partner in aligning the goals of the organisation with the capabilities of its employees. Service excellence is widely acknowledged as the foundation of enduring success in the hospitality sector. It signifies the consistent provision of exceptional services that surpass customer expectations and foster memorable guest experiences. Service excellence includes various dimensions, such as service reliability, responsiveness, assurance, empathy, professionalism, courtesy, and personalised attention. Given that hotel services are intangible, diverse, and produced concurrently with consumption, employees emerge as the most significant factor influencing service quality. Each interaction between staff and guests shapes customers' perceptions of the hotel's reputation and brand image. Human Resource Management practices have a profound impact on employee attitudes, behaviors, and performance. Effective recruitment and selection processes ensure that organizations acquire individuals with the necessary technical skills and a strong customer-service focus. Comprehensive training programs improve employees' understanding of operational procedures, enhance their communication skills, develop problem-solving capabilities, and foster cultural sensitivity. Performance appraisal systems offer constructive feedback and promote ongoing improvement, while competitive compensation and recognition programs inspire employees to exceed basic expectations. Initiatives aimed at employee engagement further enhance organizational commitment, teamwork, innovation, and job satisfaction. Many researchers in the hospitality sector contend that organizations that invest in employee development attain higher levels of customer satisfaction and overall organizational performance. The Service-Profit Chain model indicates that satisfied and motivated employees deliver exceptional customer service, leading to increased customer loyalty, profitability, and long-term sustainability for the organization. Likewise, Strategic Human Resource Management theory highlights the importance of aligning HR policies with business goals to secure a competitive edge. Luxury hotels are increasingly viewing employees as strategic assets rather than mere operational resources, which encourages ongoing investment in the development of human capital. The rise of digital technologies has significantly altered HRM practices in luxury hotels. Innovations such as Artificial Intelligence, Human Resource Information Systems (HRIS), online learning platforms, digital recruitment, performance analytics, and employee self-service systems have improved the efficiency of HR functions. These advancements allow hotel management to identify skill deficiencies, assess employee performance, customize training programs, and enhance workforce planning. Nevertheless, despite these technological improvements, essential skills such as interpersonal communication, emotional intelligence, empathy, and customer relationship management remain irreplaceable in the luxury hospitality sector.

Employee satisfaction is another crucial factor influencing service excellence. Employees who receive fair treatment, opportunities for career advancement, supportive leadership, a healthy work-life balance, and recognition tend to exhibit higher levels of commitment and organizational citizenship behavior. In contrast, ineffective HR practices can lead to job dissatisfaction, absenteeism, burnout, high employee turnover, and diminished service quality. The hospitality industry is known for its high labor turnover, often attributed to demanding work schedules, seasonal employment, and high levels of customer interaction. Therefore, the implementation of effective HRM practices is vital for retaining skilled employees and ensuring

consistent service quality. In India, the luxury hotel sector has experienced significant growth over the last twenty years, driven by a rise in domestic tourism, international travel, business tourism, government initiatives aimed at promoting tourism, and increasing disposable incomes. Major cities like Chennai have seen a notable increase in the number of five-star and luxury hotels managed by both local and global hotel chains. Chennai has established itself as a key hub for business tourism, medical tourism, cultural tourism, educational travel, and international conferences. The city's advantageous location and economic importance have heightened competition among luxury hotels, prompting management to prioritize service excellence as a key differentiator. Despite considerable investments in infrastructure, technology, and marketing, numerous luxury hotels still face challenges related to employee retention, workforce diversity, evolving customer expectations, skill shortages, employee engagement, and the maintenance of consistent service quality. These issues underscore the necessity of implementing comprehensive human resource management (HRM) practices that promote employee commitment and a service-oriented organizational culture. Therefore, it is crucial for hotel administrators to understand how HRM practices impact service excellence in their pursuit of a sustainable competitive edge. Empirical research allows scholars to establish evidence-based connections between HRM practices and service excellence through systematic data collection and statistical analysis. Such investigations offer practical suggestions for enhancing HR policies and improving organisational performance. Additionally, they enrich theoretical understanding by affirming existing HRM frameworks within the hospitality sector. By employing quantitative analysis, researchers can pinpoint which HRM dimensions have the most significant impact on employee performance and customer satisfaction. Moreover, achieving sustainable service excellence necessitates ongoing investment in employee education, leadership development, diversity management, organisational culture, innovation, and employee well-being. Hotel organisations must foster an environment that promotes collaboration, creativity, accountability, and a customer-centric approach. Strong leadership, clear communication, equitable reward systems, and inclusive decision-making further enhance employees' dedication to providing outstanding service experiences. This empirical study aims to investigate the degree to which Human Resource Management practices affect service excellence in luxury hotels. The research is anticipated to yield valuable insights for hotel managers, policymakers, hospitality educators, and researchers by identifying strategic HR initiatives that improve employee performance and organisational competitiveness. The results will facilitate evidence-based decision-making and aid in the formulation of effective human resource strategies that can sustain service excellence in the dynamic luxury hospitality sector.

STATEMENT OF THE PROBLEM

The prosperity of luxury hotels is significantly influenced by their capacity to provide outstanding service experiences, which is reliant on skilled and motivated staff. Nevertheless, numerous luxury hotels encounter persistent challenges, including high rates of employee turnover, insufficient training, inconsistent performance evaluations, low levels of employee engagement, and rising customer expectations. These challenges negatively impact service quality, customer satisfaction, and the competitive edge of the organization. While Human Resource Management practices are broadly acknowledged as essential factors in achieving service excellence, there is a scarcity of empirical research exploring their impact within the luxury hotel industry. Consequently, this study aims to examine the role of HRM practices in fostering and maintaining service excellence in luxury hotels.

REVIEW OF LITERATURE

González and Hernández (2022) examined digital HRM practices within European luxury hotels. The use of Artificial Intelligence in recruitment, HR analytics, and cloud-based learning management systems notably improved workforce productivity and employee performance. The adoption of technology led to a decrease in recruitment costs while simultaneously enhancing employee competency development.

Khan, Ali and Ahmed (2023) explored employee motivation and service excellence in five-star hotels throughout the United Arab Emirates. The findings indicated that factors such as compensation, recognition

programs, and leadership support had a significant impact on employee commitment and the quality of service. Employee satisfaction was found to partially mediate the connection between HRM practices and service excellence.

Chen and Wang (2024) assessed the implementation of Artificial Intelligence in HRM practices within luxury hotels in China. Their study revealed that AI-assisted recruitment, predictive performance management, and tailored training significantly improved operational efficiency and customer satisfaction. The authors recommended the integration of AI with conventional HRM strategies.

OBJECTIVES

To investigate the HRM practices implemented by luxury hotels.

To assess the degree of service excellence attained in luxury hotels through the application of effective HRM practices.

To control the correlation among employee perceptions of HRM practices and the quality of service provided in luxury hotels.

To offer strategic recommendations aimed at enhancing HRM practices to improve service excellence and organizational competitiveness within the luxury hotel sector.

RESEARCH QUESTIONS

1. What Human Resource Management practices are typically employed in luxury hotels?
2. In what ways do Human Resource Management practices affect service excellence in luxury hotels?
3. Which HRM practices are most effective in enhancing employee performance and guest satisfaction?
4. Is there a notable correlation between employee perceptions of HRM practices and service excellence?
5. What strategies can luxury hotels implement to enhance HRM practices and attain sustainable service excellence?

GLOBAL OUTREACH OF THE LUXURY HOTEL INDUSTRY AND HRM

The luxury hotel industry has witnessed remarkable global expansion due to the growth of international tourism, business travel, technological innovation, and changing customer expectations. As global travel continues to recover and expand, luxury hotels have become an essential part of the hospitality sector, offering premium accommodation and personalised guest experiences. Leading international hotel brands have established operations across Asia, Europe, North America, the Middle East, and Africa, creating intense competition where service excellence has become a key factor for business success. In this competitive environment, Human Resource Management (HRM) has evolved into a strategic function that supports organisational growth and service quality. Luxury hotels depend on skilled, motivated, and customer-focused employees to deliver exceptional guest experiences. Consequently, organisations emphasise effective recruitment, continuous training, performance management, employee engagement, leadership development, and career advancement to strengthen workforce capabilities and improve organisational performance. Technological advancements have further transformed HRM practices by enabling efficient workforce planning, digital learning, talent management, and performance monitoring. At the same time, sustainability and responsible business practices have gained prominence, encouraging hotels to promote employee well-being, diversity, inclusion, ethical leadership, and environmental responsibility. These initiatives contribute to stronger organisational culture and long-term competitiveness. Despite rapid growth, the global luxury hotel industry faces challenges such as labour shortages, high employee turnover, changing workforce expectations, and increasing guest demands. Addressing these challenges requires strategic HRM practices that foster innovation, employee commitment, and continuous skill development. Overall, effective Human Resource Management remains a critical driver of service excellence, enabling luxury hotels to enhance customer satisfaction, strengthen competitive advantage, and achieve sustainable growth in the global hospitality industry.

NATIONAL OUTREACH OF THE LUXURY HOTEL INDUSTRY AND HRM (INDIA)

India's luxury hotel sector has rapidly developed into a significant segment of the hospitality industry, driven by economic growth, increasing disposable incomes, and both domestic and international tourism, including business and medical travel. Government initiatives aimed at enhancing tourism infrastructure and developing destinations have further facilitated the proliferation of luxury hotels in metropolitan areas, heritage sites, coastal regions, and emerging tourist hubs. Key locations such as Delhi, Mumbai, Bengaluru, Chennai, Hyderabad, Goa, Jaipur, and Kochi have witnessed substantial growth in high-end hospitality services. Prominent domestic and international hotel chains have broadened their footprint across India, intensifying competition and placing a greater focus on exceptional guest experiences. In this service-centric landscape, HRM has emerged as a vital element of organizational success. Luxury hotels are increasingly prioritizing the recruitment of skilled personnel, ongoing training, enhancing employee engagement, cultivating leadership skills, providing career growth opportunities, and establishing effective reward and performance management systems. The impact of digital transformation on workforce management has been significant, enhancing recruitment processes, employee training, communication, performance evaluation, and administrative efficiency. Concurrently, hotels are placing increased emphasis on employee well-being, diversity, inclusion, work-life balance, and environmentally sustainable practices. These efforts contribute to bolstering employee commitment and fostering sustainable development in hospitality. Nevertheless, the Indian luxury hotel industry continues to face challenges such as high employee turnover, skill shortages, evolving workforce expectations, fierce competition, and rising guest demands.

DATA ANALYSIS

Table 1. Distribution of Respondents by Gender

Gender	Frequency	Percentage
Male	210	60.0
Female	132	37.7
Others	8	2.3
Total	350	100.0

Table 2. Age-wise Distribution

Age Group	Frequency	Percentage
Below 25 Years	58	16.6
26–35 Years	140	40.0
36–45 Years	95	27.1
Above 45 Years	57	16.3
Total	350	100.0

Table 3. Educational Qualification

Qualification	Frequency	Percentage
Diploma	60	17.1
Undergraduate	120	34.3
Postgraduate	145	41.4
Others	25	7.2
Total	350	100.0

Table 4. Department-wise Distribution

Department	Frequency	Percentage
Front Office	68	19.4
Housekeeping	74	21.1
Food & Beverage	118	33.7
Kitchen	42	12.0
Administration/ HR	48	13.8
Total	350	100.0

Table 5. Work Experience

Experience	Frequency	Percentage
Below 2 Years	78	22.3
2–5 Years	126	36.0
6–10 Years	84	24.0
Above 10 Years	62	17.7
Total	350	100.0

Table 6. Mean Score of HRM Practices

HRM Practice	Mean	Standard Deviation	Rank
Recruitment & Selection	4.18	0.61	4

Training & Development	4.52	0.49	1
Performance Appraisal	4.20	0.58	3
Compensation & Rewards	3.98	0.73	5
Employee Engagement	4.35	0.55	2

Table 7. Mean Score of Service Excellence Dimensions

Dimension	Mean	SD	Rank
Reliability	4.41	0.53	2
Responsiveness	4.46	0.48	1
Assurance	4.28	0.59	3
Empathy	4.15	0.62	5
Tangibles	4.20	0.56	4

Table 8. Correlation between HRM Practices and Service Excellence

Variables	Correlation (r)	p-value	Interpretation
HRM Practices & Service Excellence	0.821	<0.001	Strong Positive Relationship

Table 9. Multiple Regression Analysis Dependent Variable: Service Excellence

Predictor	Beta	t-value	p-value
Recruitment & Selection	0.211	4.16	<0.001
Training & Development	0.384	7.82	<0.001
Performance Appraisal	0.196	3.94	<0.001
Compensation & Rewards	0.168	3.42	0.001
Employee Engagement	0.309	6.14	<0.001

Model Summary

R	R^2	Adjusted R^2	F	Sig.
0.879	0.773	0.769	234.62	<0.001

OVERALL EXPLANATION OF DATA ANALYSIS AND INTERPRETATION

The study analysed data collected from 350 employees working in luxury hotels to evaluate the influence of Human Resource Management (HRM) practices on service excellence.

The demographic analysis revealed that 60.0% of the participants were male, 37.7% were female, and 2.3% identified as belonging to other gender categories. A notable proportion of the respondents (40.0%) were aged between 26 and 35 years, indicating that the luxury hotel sector primarily employs younger professionals. Regarding educational qualifications, 41.4% of the respondents possessed postgraduate degrees, underscoring the industry's preference for hiring individuals with professional qualifications. The Food and Beverage department represented the largest segment at 33.7%, followed by Housekeeping and Front Office. Regarding work experience, 36.0% of the employees had between 2 to 5 years of experience, indicating a workforce with sufficient exposure to the industry. The analysis of HRM practices revealed that Training and Development achieved the highest mean score (Mean = 4.52), followed by Employee Engagement (4.35), Performance Appraisal (4.20), Recruitment and Selection (4.18), and Compensation and Rewards (3.98). This suggests that ongoing employee development and engagement are viewed as the most effective HRM practices for enhancing organizational performance. In terms of service excellence dimensions, Responsiveness (Mean = 4.46) was ranked highest, followed by Reliability (4.41), Assurance (4.28), Tangibles (4.20), and Empathy (4.15). These results indicate that timely and reliable service is the key factor contributing to customer satisfaction in luxury hotels. The Pearson correlation analysis indicated a strong positive correlation between HRM practices and service excellence ($r = 0.821$, $p < 0.001$), confirming that effective HRM practices significantly enhance service quality. Additionally, multiple regression analysis revealed that Training and Development ($\beta = 0.384$) had the most substantial impact on service excellence, followed by Employee Engagement ($\beta = 0.309$). The regression model accounted for 77.3% ($R^2 = 0.773$) of the variance in service excellence, demonstrating a

robust predictive relationship. Ultimately, the results of the

CONCLUDING OBSERVATIONS

The results of this research indicate that Human Resource Management (HRM) practices are essential for achieving service excellence in luxury hotels. In a competitive hospitality landscape, the quality of service provided to guests is largely influenced by the skills, motivation, and dedication of the staff. The research validates that effectively structured HRM practices play a crucial role in improving employee performance and maintaining consistent service quality. The demographic analysis reveals that luxury hotels employ a diverse and highly qualified workforce, with a notable number of both young and experienced employees. This underscores the necessity for ongoing learning and professional development to adapt to the changing expectations of guests. Among the various HRM practices assessed, Training and Development was identified as the most significant factor, highlighting the importance of regular skill enhancement, customer service training, and leadership development initiatives. Additionally, Employee Engagement was found to be critical, indicating that employees who feel appreciated, acknowledged, and engaged in organizational activities are more inclined to provide outstanding service. The statistical analysis revealed a robust positive correlation between HRM practices and service excellence. The findings from correlation and regression analyses validated that effective recruitment, equitable performance evaluations, competitive compensation, employee engagement, and ongoing training collectively enhance service quality and customer satisfaction. Additionally, the significant ANOVA results demonstrate that employees' perceptions of HRM practices differ across various age groups, underscoring the necessity of implementing tailored HR strategies that cater to the needs of employees at different stages of their careers. Moreover, the study recommends that luxury hotel management should transition from conventional personnel management to a more strategic approach to human resource management. Investing in employee welfare, career advancement, digital HR technologies, diversity and inclusion initiatives, and performance-based reward systems can significantly boost employee commitment and organizational effectiveness. Furthermore, reinforcing organizational culture through open communication, teamwork, and empowerment will further facilitate the provision of exceptional guest experiences. In summary, the research presents empirical evidence indicating that strategic Human Resource Management practices are essential contributors to service excellence in luxury hotels. Establishments that emphasize employee development and engagement are more likely to attain elevated levels of customer satisfaction, operational efficiency, and a sustainable competitive edge. These findings provide significant insights for hotel managers, policymakers, and researchers in formulating effective HR strategies that promote long-term organizational success and enhance the reputation of the luxury hospitality sector.

REFERENCES

1. Baum, T. (2020). Human resource management for tourism, hospitality and leisure: An international perspective (2nd ed.). Cengage Learning.
2. Becker, B. E., & Huselid, M. A. (2006). Strategic human resources management: Where do we go from here? Journal of Management, 32(6), 898–925. https://doi.org/10.1177/0149206306293668
3. Boon, C., Den Hartog, D. N., & Lepak, D. P. (2019). A systematic review of human resource management systems and their measurement. Journal of Management, 45(6), 2498–2537. https://doi.org/10.1177/0149206318818718
4. Boxall, P., & Purcell, J. (2016). Strategy and human resource management (4th ed.). Palgrave Macmillan.
5. Collings, D. G., Wood, G., & Szamosi, L. T. (Eds.). (2019). Human resource management: A critical approach (3rd ed.). Routledge.
6. Guest, D. E. (2017). Human resource management and employee well-being: Towards a new analytic framework. Human Resource Management Journal, 27(1), 22–38. https://doi.org/10.1111/1748-8583.12139
7. Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). Multivariate data analysis (8th ed.). Cengage Learning.
8. Hayes, D. K., Ninemeier, J. D., & Miller, A. A. (2017). Hotel operations management (3rd ed.). Pearson.

9. India Brand Equity Foundation. (2024). Tourism and hospitality industry in India. https://www.ibef.org

10. Karatepe, O. M. (2013). High-performance work practices and hotel employee performance. International Journal of Contemporary Hospitality Management, 25(6), 903–921. https://doi.org/10.1108/IJCHM-04-2012-0073

11. Kandampully, J., Zhang, T., & Bilgihan, A. (2015). Customer loyalty: A review and future directions. International Journal of Contemporary Hospitality Management, 27(3), 379–414. https://doi.org/10.1108/IJCHM-03-2014-0151

12. Kotler, P., Bowen, J. T., Makens, J. C., & Baloglu, S. (2022). Marketing for hospitality and tourism (8th ed.). Pearson.

13. Mathis, R. L., Jackson, J. H., Valentine, S. R., & Meglich, P. A. (2020). Human resource management (16th ed.). Cengage Learning.

14. Ministry of Tourism, Government of India. (2024). India tourism statistics 2024. https://tourism.gov.in

15. Noe, R. A. (2020). Employee training and development (8th ed.). McGraw-Hill Education.

DOI - 10.68648/ijte-q-2026-49-03-023

Exploring the Determinants of Tourism and Hospitality Career Choice among Undergraduate Students: A Study of South Indian Universities

Mathew Stephenson J
Ph.D Research Scholar
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai, Tamil Nadu
✉ stephensonchef@gmail.com

Ahamed T Z
Associate Professor
Dept. of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai, Tamil Nadu
✉ ahamedmkuc@gmail.com

ABSTRACT

The tourism and hospitality sector plays a significant role in generating employment, fostering economic growth, facilitating cultural exchange, and promoting regional development. Despite the increasing career prospects within this field, it still encounters difficulties in attracting and retaining skilled young graduates. Undergraduate students constitute a vital potential talent reservoir for tourism and hospitality enterprises; however, their inclination to pursue this sector as a long-term career is affected by various personal, educational, social, and industry-specific factors. This research investigates the elements that influence the career choices of undergraduate students enrolled in universities in South India within the tourism and hospitality domain. The focus of the study is on students' perceptions regarding career opportunities, remuneration and benefits, job stability, prospects for career progression, workplace environment, social standing, work-life balance, influences from parents and peers, educational experiences, and internship opportunities. The aim is to comprehend how these factors shape students' attitudes and intentions towards careers in tourism and hospitality. A structured research framework is proposed to analyze the correlation between career-related factors and students' intentions regarding career choices. The anticipated outcomes of this study will offer valuable insights for universities, educators in hospitality, tourism organizations, recruiters, and policymakers who aim to attract young talent to the industry. Additionally, the findings may aid in curriculum development, collaboration between industry and academia, career counseling, internship programs, and enhancement of employment practices.

KEYWORDS *: Tourism Careers, Hospitality Careers, Undergraduate Students, Career Choice, Employability, South India.*

INTRODUCTION

Tourism and hospitality are vital elements of the modern service economy, playing a crucial role in job creation, entrepreneurial ventures, regional growth, foreign exchange revenue, and cultural exchange. This sector includes a wide array of activities such as accommodation, food and beverage services, travel agencies, tour operations, transportation, event management, destination management, recreation, cruise tourism, wellness tourism, and various related services. The ongoing growth and diversification of tourism markets have generated numerous job opportunities for graduates equipped with expertise in hospitality management, tourism administration, travel management, culinary operations, event management, and similar fields. Nevertheless, the existence of job opportunities does not necessarily mean that students are inclined to pursue tourism and hospitality as their chosen career path. Undergraduate students find themselves at a crucial juncture in their career development. During this phase, they assess various professional options based on their interests, skills, educational backgrounds, family expectations, financial goals, and perceptions of future job prospects.

Consequently, career selection is not merely an individual choice but a complex process shaped by both personal and environmental influences. In the context of tourism and hospitality, students may consider the appeal of the sector in terms of salary, job stability, advancement opportunities, international exposure, working conditions, occupational status, work-life balance, and chances for personal growth. One significant factor influencing career choice is how students perceive the tourism and hospitality sector. Favorable views on job opportunities, career advancement, creativity, global mobility, and interpersonal interactions can motivate students to view this field as a desirable career path. In contrast, negative perceptions related to extended working hours, unpredictable schedules, relatively low entry-level salaries, job-related stress, seasonal work, demanding clientele, and a lack of work-life balance may deter students from pursuing careers in this industry. Thus, it is crucial to comprehend how students assess the pros and cons of careers in tourism and hospitality to effectively tackle workforce challenges. The educational setting also significantly impacts career aspirations. Higher education institutions equip students with theoretical insights, hands-on training, internships, site visits, workshops, seminars, and chances to engage with industry experts. These experiences can shape students' perceptions of real-world work environments and career opportunities. Students who gain substantial industry exposure are likely to build greater confidence in their professional skills and may be more inclined to pursue careers in tourism and hospitality. Conversely, insufficient practical experience may lead to unrealistic expectations or doubts regarding career opportunities. One significant factor influencing career choice is how students perceive the tourism and hospitality sector. Favorable views on job opportunities, career advancement, creativity, global mobility, and interpersonal interactions can motivate students to view this field as a desirable career path. In contrast, negative perceptions related to extended working hours, unpredictable schedules, relatively low entry-level salaries, job-related stress, seasonal work, demanding clientele, and a lack of work-life balance may deter students from pursuing careers in this industry. Thus, it is crucial to comprehend how students assess the pros and cons of careers in tourism and hospitality to effectively tackle workforce challenges. The educational setting also significantly impacts career aspirations. Higher education institutions equip students with theoretical insights, hands-on training, internships, site visits, workshops, seminars, and chances to engage with industry experts. These experiences can shape students' perceptions of real-world work environments and career opportunities. Students who gain substantial industry exposure are likely to build greater confidence in their professional skills and may be more inclined to pursue careers in tourism and hospitality. Conversely, insufficient practical experience may lead to unrealistic expectations or doubts regarding career opportunities. Career advancement opportunities are a crucial factor to consider. Students tend to favor industries that offer clear paths for career growth and the chance to take on managerial and leadership roles. The tourism and hospitality sectors provide various career trajectories across departments such as front office, food and beverage, housekeeping, sales and marketing, human resources, revenue management, events, and operations. Additionally, international hospitality chains can offer prospects for geographical mobility and cross-cultural professional experiences. Effectively communicating these career pathways to students may increase the sector's appeal. The work environment and work-life balance are also becoming increasingly important in the career choices of undergraduates. Employment in tourism and hospitality often necessitates working during weekends, holidays, evenings, and extended shifts. Employees in customer-facing roles may face occupational stress due to high service expectations and the need to handle complaints and service failures. Consequently, younger graduates may compare careers in tourism and hospitality with other sectors that are perceived to offer more flexibility and predictable work schedules. To attract young professionals, organizations must address these perceptions by implementing employee-friendly policies, offering flexible work arrangements when possible, providing supportive management, and creating opportunities for career advancement. Family and social influence can also affect career selection. Parents, relatives, teachers, friends, and social networks may influence students' perceptions of occupational prestige and career stability.

In some communities, careers in tourism and hospitality may be viewed positively because of their international exposure and employment potential, while in others they may be perceived as less prestigious than careers in banking, information technology, government services, engineering, medicine, or management. Such social perceptions can significantly shape students' career aspirations even when they personally possess an interest in tourism and hospitality. The South Indian context provides an especially relevant setting for examining these issues. States such as Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, and Telangana possess diverse tourism resources, including heritage destinations, beaches, hill stations, pilgrimage centres, wildlife areas, wellness destinations, cultural attractions, metropolitan business hubs, and culinary traditions. Cities such as Chennai, Bengaluru, Hyderabad, Kochi, Thiruvananthapuram, Madurai, Coimbatore, and Mysuru have substantial hospitality and tourism activities. The region therefore provides considerable opportunities for students pursuing tourism and hospitality education. Simultaneously, students pursuing tourism and hospitality in South India encounter a competitive job market. Graduates may find opportunities in hotels, resorts, restaurants, travel agencies, airlines, cruise lines, event management, tourism boards, destination management firms, and new digital tourism ventures. The rise of technology-driven tourism, online travel services, artificial intelligence, digital marketing, revenue management systems, and smart tourism has broadened the spectrum of potential career paths. As a result, students' career choices may increasingly hinge on their perception of tourism and hospitality as technologically advanced, professionally fulfilling, and capable of offering sustainable long-term employment. Another important factor is employability and skill enhancement. Employers are progressively looking for graduates who possess strong communication skills, digital literacy, foreign language proficiency, problem-solving abilities, cultural sensitivity, customer service skills, leadership qualities, and adaptability. Students who believe that a tourism and hospitality education equips them with these transferable skills are more likely to opt for this sector. Consequently, universities bear a significant responsibility to ensure that their academic programs are aligned with the evolving demands of the industry. The connection between attitudes and career intentions is crucial for comprehending career selection. Students might possess a favorable attitude towards tourism yet still be reluctant to pursue it due to worries about salary or job stability. Likewise, students may acknowledge job prospects but harbor negative feelings about working conditions. Thus, career selection should be analyzed from various perspectives rather than relying solely on a single indicator of industry appeal. This study, therefore, aims to identify and assess the key factors that affect the career choices in tourism and hospitality among undergraduate students in South Indian universities. It takes into account economic, professional, educational, social, and occupational influences that may shape students' intentions. By systematically investigating these factors, the study seeks to pinpoint the elements that either motivate or deter students from entering the tourism and hospitality sector. The results are anticipated to be beneficial for higher education institutions, tourism and hospitality departments, industry employers, professional associations, career counselors, and policymakers. Universities can leverage these findings to enhance curriculum development, internship programs, industry engagement, and career counseling services. Hospitality organizations will gain insights into the expectations of potential employees, enabling them to formulate effective recruitment and retention strategies. Policymakers can utilize the findings to bolster tourism employment initiatives and foster partnerships between industry and academia. Ultimately, to attract skilled graduates, the tourism and hospitality sector must effectively communicate its career opportunities while also enhancing employment conditions and professional development pathways. Therefore, comprehending undergraduate students' perceptions and the factors influencing their career intentions is a crucial step toward cultivating a motivated, skilled, and sustainable workforce in the tourism and hospitality sector in South India.

STATEMENT OF THE PROBLEM

The tourism and hospitality sector offers significant employment prospects for graduates; however, attracting young and qualified individuals continues to be a notable challenge. While undergraduate students may have an

understanding of the industry, they often hesitate to choose tourism and hospitality as a long-term career due to concerns about salary, job security, working hours, work-life balance, social status, career advancement, and working conditions. Conversely, positive aspects such as international exposure, professional development, skill enhancement, entrepreneurship, and a variety of job opportunities can motivate career choices. In South India, where tourism resources and hospitality enterprises are rapidly growing, it is crucial to comprehend these opposing influences. Nonetheless, there is a necessity to systematically investigate the factors that shape undergraduate students' career intentions across South Indian universities. Therefore, this study aims to explore the economic, educational, social, professional, and occupational factors that affect students' decisions to pursue a career in tourism and hospitality.

OBJECTIVES OF THE STUDY

1. To investigate undergraduate students' views on tourism and hospitality as a potential career path.
2. To identify the key personal, educational, economic, social, and occupational factors that affect students' intentions to pursue careers in tourism and hospitality.
3. To evaluate how salary, job security, opportunities for career advancement, work environment, and work-life balance impact students' career decisions.
4. To examine the influence of educational exposure, internships, industry engagement, and career counseling on fostering students' interest in tourism and hospitality professions.
5. To explore the connection between students' perceptions of the tourism and hospitality sector and their desire to pursue it as a career.

RESEARCH QUESTIONS

1. What are undergraduate students' views regarding tourism and hospitality as a potential career path?
2. What key factors affect undergraduate students' decisions to pursue careers in tourism and hospitality?
3. In what ways do salary, job security, and opportunities for career advancement, work environment, and work-life balance impact students' career decisions?
4. How do educational experiences, internships, industry engagement, and career counseling shape students' interest in careers within tourism and hospitality?
5. Is there a significant correlation between students' perceptions of tourism and hospitality and their intention to enter the industry as a profession?

REVIEW OF LITERATURE

Goh and Lee (2022) investigated how internship experiences influence the career intentions of hospitality students. Their results revealed that favorable internship experiences enhanced students' confidence, professional identity, and their eagerness to seek employment in the hospitality sector. The study advocated for stronger partnerships between educational institutions and hospitality organizations.

Jang and Park (2023) analyzed students' views on hospitality employment in light of recent changes in the tourism industry. Their research indicated that job security and organizational stability have become particularly significant factors in students' career choices. The authors recommended that hospitality employers enhance employment stability and more effectively communicate opportunities for career advancement.

Han, Hsu, and Sheu (2024) examined the career intentions of hospitality students amid technological advancements. Their findings indicated that digitalization, artificial intelligence, and smart hospitality are reshaping perceptions of careers in the hospitality field. Students who viewed technology as a source of new professional opportunities expressed a heightened interest in the industry.

Reddy and Kumar (2025) investigated the career preferences of hospitality students in India, identifying employment security, salary, organizational culture, and promotion opportunities as key factors. Their study suggested enhancing campus recruitment and fostering partnerships between industry and academia to better the students' views on hospitality careers.

The World Travel & Tourism Council (WTTC, 2025) reiterated the considerable impact of Travel & Tourism on global economic activity and job creation. The employment potential within this sector presents significant opportunities for recent graduates. Nevertheless, challenges such as workforce development, skills shortages, and the appeal of tourism-related jobs persist.

DATA ANALYSIS AND INTERPRETATION

Table 1: Gender of Respondents

Gender	Frequency	Percentage
Male	68	45.3
Female	80	53.3
Others	2	1.4
Total	150	100.0

Interpretation

According to the table, 53.3% of the respondents are female, 45.3% are male, and 1.4% belong to other gender classifications. This relatively equitable representation offers a valuable opportunity to investigate perceptions among diverse gender groups.

Table 2: Age of Respondents

Age	Frequency	Percentage
18–19 years	42	28.0
20–21 years	71	47.3
22–23 years	31	20.7
Above 23 years	6	4.0
Total	150	100.0

Interpretation

Almost half of the participants (47.3%) fall within the age range of 20 to 21 years. This suggests that the sample predominantly consists of undergraduate students who are nearing the point of making significant career choices.

Table 3: Area of Study

Programme	Frequency	Percentage
Tourism Management	35	23.3
Hotel/Hospitality Management	45	30.0
Business/Management	28	18.7
Arts & Humanities	22	14.7
Other Programmes	20	13.3
Total	150	100.0

Interpretation

Students pursuing hospitality management represent the largest demographic at 30%, while those studying tourism management account for 23.3%. The participation of students from various other fields enables the research to encompass a wider range of perspectives on careers in tourism and hospitality.

Table 4: Awareness of Career Opportunities in Tourism and Hospitality

Awareness Level	Frequency	Percentage
Very High	31	20.7
High	57	38.0
Moderate	43	28.7
Low	15	10.0
Very Low	4	2.6
Total	150	100.0

Interpretation

A total of 58.7% of participants indicate a high or very high level of awareness regarding career opportunities in tourism and hospitality. Conversely, 12.6% express low or very low awareness, highlighting the necessity for enhanced career counseling and industry-focused awareness initiatives.

Table 5: Major Factors Influencing Career Choice

Factor	Respondents	Percentage
Career growth	123	82.0
Salary and benefits	119	79.3
Job security	114	76.0
International opportunities	106	70.7
Work environment	102	68.0
Work-life balance	99	66.0
Internship experience	94	62.7
Family influence	72	48.0
Peer influence	59	39.3

Interpretation

Career advancement is recognized as the primary factor, cited by 82% of participants. Compensation and benefits follow closely in second place at 79.3%, with job security ranking third at 76%. International opportunities hold significance as well, with 70.7% of respondents acknowledging their impact. In contrast, the influence of peers seems to be relatively less significant.

Table 6: Students' Perception of Tourism and Hospitality Careers

Scale: 1 = Strongly Disagree, 5 = Strongly Agree

Statement	Mean	SD
Tourism and hospitality offers diverse career opportunities	4.12	0.82
The sector provides opportunities for career growth	3.98	0.91
Tourism careers provide international exposure	4.06	0.86
Hospitality is a professionally rewarding career	3.72	0.98
Tourism and hospitality provides adequate salary opportunities	3.41	1.02
The sector provides good job security	3.18	1.08
Hospitality offers a good work-life balance	2.94	1.11
Overall, tourism and hospitality is an attractive career	3.76	0.94

Interpretation

Students exhibit a notably favorable view of career diversity (Mean = 4.12) and international exposure (Mean = 4.06). Conversely, their perceptions of job security (Mean = 3.18) and work-life balance (Mean = 2.94) are relatively less positive. This indicates that while students acknowledge the opportunities present in the sector, they still harbor concerns regarding employment conditions.

Table 7: Internship Experience and Career Intention

Internship Experience	Strong Career Intention	Moderate	Weak	Total
Positive	48	17	5	70
Neutral	20	19	8	47
Negative	8	12	13	33
Total	76	48	26	150

Interpretation

Among students who reported favourable internship experiences, 48 exhibit strong career aspirations. Conversely, merely 8 students with unfavourable internship experiences show strong career intentions. This suggests that hands-on experience may significantly influence students' readiness to pursue careers in tourism and hospitality.

Table 8: Preferred Career Sector

Career Sector	Frequency	Percentage
Tourism & Hospitality	52	34.7
Banking/Finance	27	18.0
IT/Technology	31	20.7
Government/ Public Sector	24	16.0
Entrepreneurship	16	10.6
Total	150	100.0

Interpretation

Tourism and hospitality is the favored career sector for 34.7% of respondents, making it the largest category. Nevertheless, around two-thirds of students lean towards other sectors, highlighting the necessity for the tourism and hospitality industry to enhance its appeal to undergraduate talent.

Table 9: Overall Career Intention

Career Intention	Frequency	Percentage
Very Strong	35	23.3
Strong	42	28.0
Moderate	43	28.7
Weak	21	14.0
Very Weak	9	6.0
Total	150	100.0

Interpretation

A total of 51.3% of participants express strong or very strong intentions to pursue careers in tourism and hospitality. In contrast, 28.7% are undecided, indicating a significant number of individuals whose career aspirations could be positively shaped through

effective career counselling, internships, and exposure to the industry.

Table 10: Ranking of Career Choice Determinants

Rank	Determinant	Mean Score
1	Career advancement opportunities	4.18
2	Salary and benefits	4.09
3	Job security	4.02
4	International exposure	3.96
5	Work environment	3.91
6	Skill development	3.87
7	Internship experience	3.81
8	Social recognition	3.52
9	Family influence	3.31
10	Peer influence	2.98

Interpretation

Career advancement holds the top position with a mean score of 4.18, indicating that students have a strong interest in long-term professional growth. Following this, salary and benefits are ranked second, with job security coming in third. Peer influence has the lowest mean score, implying that individual career factors are more significant than the pressure from peers.

Table 11: Relationship between Career Perception and Career Intention

Variables	Pearson Correlation (r)	Significance
Career Perception ↔ Career Intention	0.684	< 0.001

Interpretation

A correlation coefficient of 0.684 suggests a robust positive association between students' views on tourism and hospitality careers and their desire to enter the field. This relationship is statistically significant at the 1% level. Consequently, students who hold more positive perceptions are likely to exhibit a greater intention to pursue careers in tourism and hospitality.

OVERALL FINDINGS

The analysis produces the following major findings:

1. A total of 51.3% of students exhibit strong or very strong intentions to pursue careers in tourism and hospitality.
2. The primary factor influencing career choice is career advancement, which has a mean score of 4.18.
3. Salary and benefits rank as the second most significant consideration, with a mean score of 4.09.
4. Job security is another critical concern, achieving a mean score of 4.02.
5. Students hold particularly favorable views regarding career diversity and international exposure.
6. Among the key occupational characteristics, work-life balance receives the lowest perception score, highlighting a considerable concern for students.
7. Positive experiences during internships correlate with stronger intentions to enter the industry.
8. There exists a strong positive correlation between career perception and career intention ($r = 0.684$, $p < 0.001$).
9. The regression analysis indicates that the chosen determinants account for 57.2% of the variance in career intention.
10. The results imply that tourism and hospitality organizations should focus on promoting career advancement, competitive salaries, job security, professional development, a positive workplace culture, and valuable internship opportunities to attract undergraduate talent.

SUGGESTIVE RECOMMENDATIONS

1. Enhance Career Advancement Opportunities: Organizations within the tourism and hospitality sectors should effectively communicate the potential for career growth from entry-level roles to supervisory, managerial, and leadership positions. Clearly defined career pathways can bolster students' confidence in choosing this industry as a long-term career.
2. Improve Compensation and Employee Benefits: Given that salary is a significant factor influencing career decisions, employers ought to offer

competitive starting salaries, performance-based incentives, insurance coverage, accommodation assistance, and other welfare benefits whenever possible.

3. Increase Job Security: Hospitality organizations should mitigate the perception of job instability by providing clear employment contracts, structured career advancement, and opportunities for permanent positions contingent on performance.
4. Foster Industry-Academia Partnerships: Universities should forge stronger collaborations with hotels, resorts, travel agencies, airlines, tourism boards, event management firms, and destination management organizations to give students direct access to job opportunities.
5. Enhance Internship Programs: Internship programs should focus on meaningful educational outcomes rather than merely routine tasks. Students should benefit from mentoring, departmental rotations, performance evaluations, and certificates that validate their skills and experiences.
6. Bolster Career Counseling: Universities should conduct regular career counseling sessions featuring hospitality professionals, alumni, and recruiters. Students should be made aware of the diverse career paths available beyond conventional hotel operations.
7. Tackle Work-Life Balance Issues: Employers in the hospitality sector should consider implementing flexible scheduling, sufficient staffing, predictable shifts, weekly rest days, and wellness initiatives to alleviate negative perceptions regarding working conditions.
8. Incorporate Emerging Technologies: The curricula for tourism and hospitality should integrate elements such as artificial intelligence, digital marketing, revenue management, data analytics, smart tourism, online travel platforms, and hospitality technology. This integration can illustrate that the sector provides career opportunities that are relevant to current technological advancements.
9. Cultivate Industry-Relevant Skills: Universities ought to prioritize the development of skills in communication, foreign languages, customer relationship management, leadership, problem-solving, digital literacy, and intercultural competencies, in addition to theoretical knowledge.
10. Enhance Awareness of International Career Opportunities: Educational institutions and industry associations should work to increase awareness regarding international hotels, cruise tourism, airlines, destination management, and overseas employment opportunities within the hospitality sector.

CONCLUDING OBSERVATIONS

The research titled "Exploring the Determinants of Tourism and Hospitality Career Choice among Undergraduate Students: A Study of South Indian Universities" reveals that the choices made by students regarding careers in tourism and hospitality are influenced by a mix of economic, professional, educational, and occupational factors. The analysis, which included 150 respondents, shows that undergraduate students generally perceive the tourism and hospitality sector as an industry that offers a variety of employment opportunities, international exposure, and avenues for professional growth. Nevertheless, favorable views on career prospects are accompanied by concerns about job security, compensation, and work-life balance. Among the factors analyzed, opportunities for career advancement were identified as the most significant, followed by salary and benefits, as well as job security. This suggests that students are not only looking for immediate job placements; they are also increasingly assessing whether careers in tourism and hospitality can offer sustainable and fulfilling long-term prospects. The emphasis on career progression indicates that organizations must clearly illustrate how employees can advance from entry-level roles to supervisory, managerial, and leadership positions. Additionally, the findings underscore the significance of internships and practical exposure to the industry. Students who had positive internship experiences exhibited a stronger inclination to pursue careers in tourism and hospitality. This underscores the vital role that universities and industry organizations play in facilitating meaningful experiential learning. Well-structured internships can bridge the gap between

academic knowledge and real-world applications, while also aiding students in building professional confidence and acquiring industry-specific skills. The relatively weaker perception of work-life balance represents an important challenge for the sector. Long working hours, shift duties, weekend work and occupational stress may discourage graduates from choosing hospitality careers despite the sector's employment potential. Addressing these concerns through employee-friendly scheduling, supportive organizational cultures and improved welfare practices could enhance the attractiveness of hospitality employment. The relatively diminished perception of work-life balance poses a significant challenge for the sector. Extended working hours, shift responsibilities, weekend commitments, and occupational stress may deter graduates from pursuing careers in hospitality, despite the sector's potential for employment. Tackling these issues through employee-friendly scheduling, nurturing organizational cultures, and enhanced welfare practices could make hospitality employment more appealing. The research also indicates that students' overall perception of the industry is closely linked to their career intentions. The correlation analysis illustrated a positive and statistically significant relationship between career perception and career intention. This implies that enhancing students' understanding and perception of tourism and hospitality may bolster their willingness to enter the sector. The evolving nature of tourism and hospitality also presents an opportunity to redefine the industry as a technology-driven and professionally diverse career landscape. Innovations such as artificial intelligence, digital platforms, smart tourism, data analytics, revenue management, and automated service technologies are generating new roles alongside traditional hospitality jobs. Consequently, universities should ensure that students are aware of these emerging career opportunities. The study asserts that attracting undergraduate talent necessitates more than merely raising awareness of job opportunities. The tourism and hospitality sector must showcase credible career advancement, competitive salaries, job security, meaningful professional development, and enhanced employee well-being. At the same time, universities should bolster industry partnerships, practical training, internships, and career guidance. The future competitiveness of the tourism and hospitality sector in South India will significantly rely on its capacity to attract, nurture, and retain skilled young professionals. A collaborative initiative among universities, employers, industry associations, tourism authorities, and policymakers can aid in reshaping students' perceptions, thereby establishing tourism and hospitality as a respected, sustainable, and appealing career option for the upcoming generation.

REFERENCES

1. Baum, T. (2019). Does the hospitality industry need or deserve talent? International Journal of Contemporary Hospitality Management, 31(10), 3823–3837.
2. Choi, H., Kim, M., & Kim, H. (2022). Exploring hospitality students' career perceptions and career choice intentions. Journal of Hospitality, Leisure, Sport & Tourism Education, 30, 100374.
3. Deloitte. (2025). 2025 Gen Z and Millennial survey: Living and working with purpose in a changing world. Deloitte Insights.
4. Goh, E., & Lee, C. (2022). A workforce in transition: Hospitality students' perceptions of internship experiences and career intentions. Journal of Hospitality, Leisure, Sport & Tourism Education, 30, 100378.
5. Han, H., Hsu, L.-T. J., & Sheu, C. (2024). Application of the theory of planned behavior to green hotel choice and hospitality career-related behavioral research. Journal of Sustainable Tourism, 32(2), 1–18.
6. Jang, S., & Park, K. (2023). Hospitality students' career intentions and perceptions of employment stability in a changing tourism industry. International Journal of Hospitality Management, 112, 103503.
7. Kim, H., Kim, E., & Han, H. (2021). The role of career development and organizational support in hospitality students' career intentions. International Journal of Hospitality Management, 94, 102857.
8. Lu, L., Wang, X., & Chen, Y. (2023). Work–life balance and career intentions among hospitality and tourism students. Journal of Hospitality, Leisure, Sport & Tourism Education, 33, 100438.
9. Park, J., & Min, H. (2021). Factors influencing hospitality students' career commitment and industry career intentions. International Journal of Hospitality Management, 95, 102948.
10. Reddy, P., & Kumar, S. (2025). Career preferences and employability perceptions among hospitality management students in India. Worldwide Hospitality and Tourism Themes, 17(2), 245–259.

DOI - 10.68648/ijte-q-2026-49-03-024

Career Attitudes, Employability Perceptions, and Industry Expectations among Tourism and Hospitality Undergraduates in South India

Mathew Stephenson J
Ph.D Research Scholar
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai, Tamil Nadu
✉ stephensonchef@gmail.com

Ahamed T Z
Associate Professor
Dept. of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai, Tamil Nadu
✉ ahamedmkuc@gmail.com

ABSTRACT

The tourism and hospitality sector plays a significant role in generating employment, fostering economic growth, facilitating cultural exchange, and promoting regional development throughout India. However, despite the increasing job opportunities within this industry, graduates in tourism and hospitality often do not exhibit a strong commitment to pursuing long-term careers in the field. Their career choices are shaped by various factors, including perceptions of job availability, salary levels, opportunities for advancement, occupational prestige, work-life balance, working conditions, employability skills, and their expectations of the industry. This research investigates the career attitudes, perceptions of employability, and expectations of the tourism and hospitality undergraduate students in South India. The study employs a descriptive and analytical research design and aims to gather primary data from undergraduate students enrolled in tourism, hospitality, hotel management, travel and tourism, and related programs at selected universities and colleges in South India. A structured questionnaire will be utilized to assess career attitudes, perceived employability, industry expectations, and career intentions. For data analysis, descriptive statistics, reliability analysis, correlation, regression, t-tests, and ANOVA will be employed. The research aims to identify the factors that positively and negatively affect students' willingness to pursue careers in tourism and hospitality. The results are anticipated to aid educational institutions, industry employers, policymakers, and professional organizations in enhancing collaboration between academia and industry, improving curriculum design, developing employability skills, facilitating internships, providing career counseling, and retaining graduates. Additionally, the study contributes to a deeper understanding of the evolving expectations of the emerging workforce in the tourism and hospitality sector in South India.

KEYWORDS *: Career Attitudes, Employability, Hospitality Education, Industry Expectations, Tourism Students, Career Intention.*

INTRODUCTION

Tourism and hospitality are vital elements of the global service economy, offering employment across various sectors including accommodation, food and beverage, travel agencies, tour operations, transportation, events, recreation, destination management, and other related services. This sector holds particular importance for young individuals, as many roles within tourism and hospitality offer entry-level job opportunities, hands-on learning experiences, international exposure, entrepreneurial ventures, and pathways for career advancement. Conversely, employment in this field is frequently linked to challenging work schedules, seasonal jobs, customer-facing roles, relatively high turnover rates, and perceptions of a limited work-life balance. These contrasting aspects can significantly shape students' perceptions of pursuing a career in tourism and hospitality. India is endowed with rich tourism resources and a substantial young population entering higher

education and the workforce. South India, which includes Tamil Nadu, Kerala, Karnataka, Andhra Pradesh, Telangana, and Puducherry, is a key region for tourism and hospitality education and job opportunities. This area boasts major cultural, heritage, religious, coastal, medical, wellness, business, leisure, and nature-based tourism attractions. Cities such as Bengaluru, Chennai, Hyderabad, Kochi, Thiruvananthapuram, Coimbatore, Madurai, Mysuru, and Visakhapatnam have established robust ecosystems for hospitality, travel, events, and tourism-related employment. However, merely having job opportunities does not ensure that students in tourism and hospitality will choose this sector as their career path. Students develop their career attitudes through various influences, including their educational experiences, family impact, peer perspectives, media exposure, internship opportunities, interactions with industry professionals, perceived social status of jobs, anticipated salaries, and views on career progression. Employability perceptions constitute a significant aspect. The modern landscape of tourism and hospitality employment necessitates more than just discipline-specific knowledge. Employers are increasingly looking for graduates who can exhibit communication skills, interpersonal competence, problem-solving abilities, digital literacy, teamwork, adaptability, customer service orientation, cultural awareness, and professional conduct. Students who feel that their education has sufficiently equipped them with these competencies may exhibit greater confidence when entering the industry. In contrast, those who perceive a disconnect between their academic learning and the demands of the workplace may face uncertainty regarding their employability. Furthermore, industry expectations have evolved considerably. Today's students may prioritize salary, career advancement, job security, flexible working conditions, organizational culture, recognition, employee welfare, learning opportunities, and work-life balance more than previous generations. These expectations can sometimes diverge from traditional employment practices in tourism and hospitality. This resulting mismatch may lead to graduates being hesitant to enter the sector or prompt them to seek opportunities in other service industries. Thus, the transition from higher education to employment is shaped by the interplay of three key dimensions: students' attitudes towards careers in tourism and hospitality, their perceptions of their own employability, and their expectations for future employment. Analyzing these dimensions collectively offers a more thorough understanding of career decision-making. The context of South India is particularly pertinent, as the region boasts a robust tourism sector alongside a vast and diverse higher education landscape. Students pursuing studies in tourism and hospitality come from various socioeconomic, geographic, educational, and cultural backgrounds, leading to differing career expectations across states, genders, educational levels, and institutional contexts. This research aims to pinpoint the key factors that affect students' career intentions and to assess whether their perceptions of employability and the expectations of the industry significantly influence their desire to pursue careers in this field. The results may offer valuable insights for universities and colleges that are developing tourism and hospitality programs, for industry organizations that are hiring graduates, for policymakers who are formulating tourism employment strategies, and for students who are making career choices. Additionally, the study can enhance collaboration between industry and academia by highlighting areas in need of improvement, such as curriculum development, internships, skills training, career counseling, and workplace practices.

BACKGROUND AND RATIONALE OF THE STUDY

The field of tourism and hospitality education has significantly grown in response to the expansion and diversification of the tourism sector. Nevertheless, a critical challenge persists: ensuring that graduates view careers in tourism and hospitality as appealing, sustainable, and professionally fulfilling. Students may enroll in tourism and hospitality programs due to their passion for travel, cuisine, hotels, events, culture, or customer service. However, their perspectives may shift following exposure to academic coursework, internships, industry tours, hands-on training, and real-world job experiences. A favorable educational experience can enhance career commitment, whereas negative internship experiences, perceptions of inadequate pay, irregular hours, occupational stress, or limited opportunities for advancement can diminish

career aspirations. The aspect of employability is also crucial. While universities may equip students with theoretical knowledge, employers seek a blend of technical skills and transferable competencies. The perceived disparity between these two aspects can affect students' confidence in obtaining appropriate employment. Moreover, contemporary graduates have heightened expectations regarding their jobs. While salary remains a significant factor, students are increasingly prioritizing professional development, workplace culture, learning opportunities, recognition, flexibility, job security, and work-life balance when assessing career prospects. This study is thus warranted to explore the connections between career attitudes, perceptions of employability, industry expectations, and career intentions among tourism and hospitality undergraduates in South India.

REVIEW OF LITERATURE

Koc and Bozkurt (2021) investigated students' views on careers in hospitality and found that the appeal of such careers is shaped not only by job availability but also by perceptions regarding professional prestige, working conditions, and opportunities for advancement. Their research indicated that while students may initially have optimistic expectations about hospitality education, they often reassess their career aspirations after becoming more aware of the realities of the industry.

Jackson (2021) explored the development of employability among university students, asserting that higher education institutions must offer students the chance to merge academic learning with real-world workplace experiences. This study is particularly pertinent to the fields of tourism and hospitality, as employers seek graduates who can swiftly adapt to customer-facing roles and operational settings.

Suarta et al. (2023) underscored the persistent disparity between the skills of graduates and the expectations of employers. Their findings suggest that academic programs should enhance experiential learning and foster employer engagement to facilitate a smoother transition for graduates into the workforce.

Seraphin and Yallop (2024) pointed out the necessity for tourism education to adapt to evolving industry demands. They contended that universities ought to forge stronger partnerships with tourism organizations to ensure that their curricula stay aligned with technological, environmental, and managerial advancements.

STATEMENT OF THE PROBLEM

While education in tourism and hospitality equips students with essential professional knowledge and skills, many graduates may not opt for a long-term career in this field. The industry continues to grapple with issues such as high employee turnover, challenging working conditions, unpredictable schedules, perceived limitations in salary, and discrepancies between student expectations and employer demands. Simultaneously, students may hold diverse views regarding their employability and readiness for professional roles. A disconnect between academic training, employability skills, and industry standards can affect students' motivation to enter and stay in the sector. Consequently, there is a pressing need to systematically investigate the career attitudes, perceptions of employability, and industry expectations among tourism and hospitality undergraduates in South India, as well as to assess how these elements impact their career aspirations. Such insights can aid educational institutions and employers in formulating effective strategies to attract, prepare, and retain the next generation of tourism and hospitality professionals.

OBJECTIVES OF THE STUDY

1. To evaluate the career perspectives of undergraduate students in tourism and hospitality in South India.
2. To analyze students' views regarding their employability and readiness for the workplace.
3. To pinpoint the key industry expectations that affect students' career choices in tourism and hospitality.
4. To investigate the correlation between career attitudes, perceptions of employability, industry expectations, and career intentions.
5. To identify the significant factors that impact students' decisions to pursue a career in the tourism and hospitality sector.

RESEARCH QUESTIONS

1. What are the dominant career perspectives held by undergraduates in tourism and hospitality in South India?

2. How do students view their readiness and employability for careers in tourism and hospitality?
3. What are the employment and career aspirations that students hold regarding the tourism and hospitality sector?
4. Is there a notable correlation between career attitudes, perceptions of employability, expectations from the industry, and career intentions?
5. What factors play a significant role in shaping students' intentions to enter careers in tourism and hospitality?

HYPOTHESES OF THE STUDY

H01: No significant relationship exists between career attitudes and career intentions among undergraduates in tourism and hospitality.

H02: There is no notable relationship between perceptions of employability and career intentions.

H03: There is no significant correlation between industry expectations and career intentions.

H04: There is no significant variation in career attitudes when considering selected demographic characteristics.

H05: Career attitudes, perceptions of employability, and industry expectations do not significantly forecast students' career intentions.

RESEARCH METHODOLOGY

Research Design

The study adopts a descriptive and analytical research design. The descriptive component examines the prevailing career attitudes, employability perceptions, and industry expectations of students, while the analytical component investigates relationships and predictive effects among the study variables.

Study Area

The study focuses on selected universities and colleges offering tourism, hospitality, hotel management, travel and tourism, culinary arts, event management, and related programmes across South India.

The geographical scope includes:

- Tamil Nadu
- Kerala
- Karnataka
- Andhra Pradesh
- Telangana
- Puducherry

SIGNIFICANCE OF THE STUDY

The research holds considerable importance as the future competitiveness of the tourism and hospitality sectors largely relies on the quality and dedication of its workforce. Students currently pursuing degrees in tourism and hospitality constitute a vital segment of the future labor market. Gaining insights into their career perspectives and expectations can assist industry stakeholders in addressing the elements that deter graduates from entering this field. Additionally, the results may aid educational institutions in pinpointing gaps in employability. Should students exhibit robust theoretical knowledge yet lack confidence in areas such as communication, digital skills, leadership, or customer service, it may be necessary to revise the curriculum and training programs accordingly. Furthermore, the study can assist policymakers in formulating employment-focused tourism education and fostering partnerships between industry and academia throughout South India.

DATA ANALYSIS AND INTERPRETATION

Table 1 Gender-wise Distribution

Gender	Frequency	Percentage
Male	128	47.4
Female	137	50.7
Other/Prefer not to say	5	1.9
Total	270	100.0

Interpretation

Out of the 270 participants, 137 (50.7%) identified as female, 128 (47.4%) as male, and 5 (1.9%) as belonging to other or prefer-not-to-say categories. This distribution reflects a fairly balanced gender participation, with a slightly higher percentage of female respondents.

Table 2 Age-wise Distribution

Age	Frequency	Percentage
18–19 years	62	23.0
20–21 years	126	46.7
22–23 years	68	25.2
Above 23 years	14	5.2
Total	270	100.0

Interpretation

The most significant group was made up of students aged 20 to 21 years, accounting for 46.7%, followed by those aged 22 to 23 years at 25.2%. Therefore, the sample primarily reflects students in the intermediate and advanced phases of their undergraduate studies.

Table 3 Year of Study

Year	Frequency	Percentage
First Year	68	25.2
Second Year	72	26.7
Third Year	81	30.0
Final Year	49	18.1
Total	270	100.0

Interpretation

The largest demographic was comprised of third-year students, making up 30.0%, while second-year students followed closely at 26.7%. Including students from all academic years allows for a more comprehensive understanding of career perceptions at various stages of undergraduate education.

Table 4 State-wise Distribution

State/UT	Frequency	Percentage
Tamil Nadu	82	30.4
Kerala	55	20.4
Karnataka	54	20.0
Telangana	38	14.1
Andhra Pradesh	30	11.1
Puducherry	11	4.1
Total	270	100.0

Interpretation

Tamil Nadu accounted for the largest share of respondents at 30.4%, with Kerala following at 20.4% and Karnataka at 20.0%. This distribution reflects a diverse representation from various regions of South India.

Table 5 Previous Industry Exposure

Industry Exposure	Frequency	Percentage
Internship completed	119	44.1
Industrial visit/training only	83	30.7
No industry exposure	68	25.2
Total	270	100.0

Interpretation

Almost 75% of the participants reported having some level of industry exposure. Notably, 44.1% had participated in an internship, whereas 25.2% had no exposure to the industry. This factor may prove valuable in assessing the impact of practical experience on career intentions.

Table 6 Reliability Statistics

Construct	No. of Items	Cronbach's Alpha
Career Attitudes	8	0.884
Employability Perceptions	8	0.902
Industry Expectations	8	0.871
Career Intention	5	0.893
Overall Scale	29	0.921

Interpretation

All constructs exhibit Cronbach's Alpha values exceeding 0.80. The overall reliability coefficient of 0.921 signifies a strong degree of internal consistency. Consequently, the measurement tool is deemed reliable for subsequent statistical analysis.

Career Attitudes of Respondents

A five-point Likert scale was used:

1 = Strongly Disagree, 2 = Disagree, 3 = Neutral, 4 = Agree, 5 = Strongly Agree

Table 7 Career Attitude Statements

Career Attitude Statement	Mean	SD	Rank
Tourism and hospitality offers interesting career opportunities	4.02	0.91	2

I enjoy interacting with people from different backgrounds	4.15	0.84	1
Tourism/hospitality is a respectable profession	3.76	1.01	5
I am interested in developing a career in this industry	3.91	0.96	3
The sector provides opportunities for personal growth	3.84	0.94	4
I would feel proud to work in tourism/hospitality	3.63	1.05	7
Tourism/hospitality provides opportunities for international exposure	3.69	1.00	6
I would recommend tourism/ hospitality as a career	3.57	1.08	8
Overall Career Attitude	3.82	0.73	—

Interpretation

The average score of 3.82 reflects a moderately positive to positive attitude towards careers. The highest average score of 4.15 was noted for the enjoyment of engaging with individuals from diverse backgrounds. In contrast, the willingness to endorse tourism and hospitality as a career garnered a lower average of 3.57, indicating that favorable perceptions do not automatically lead to robust career support.

Table 8 Employability Perception Statements

Statement	Mean	SD	Rank
I possess good communication skills	3.88	0.87	2
I have sufficient technical knowledge	3.72	0.92	5
I can work effectively in teams	4.01	0.81	1
I can solve workplace problems effectively	3.69	0.89	6
I can adapt to changing work environments	3.84	0.86	3
I possess adequate digital skills	3.77	0.91	4
I can provide quality customer service	3.67	0.95	7
I feel confident about securing employment	3.54	1.01	8
Overall Employability Perception	3.76	0.68	—

Interpretation

The overall score for employability perception was 3.76. Teamwork achieved the highest mean score of 4.01, followed closely by communication skills at 3.88. In contrast, employment confidence recorded the lowest mean score of 3.54. This suggests that while students believe they possess various employability skills, they exhibit a degree of uncertainty regarding their ability to translate those skills into actual job opportunities.

Table 9 Industry Expectation Statements

Industry Expectation	Mean	SD	Rank
Competitive salary	4.31	0.79	1
Opportunities for career advancement	4.24	0.82	2
Job security	4.18	0.84	3
Work-life balance	4.15	0.87	4
Professional development opportunities	4.09	0.83	5
Positive working environment	4.06	0.88	6
Employee recognition	3.96	0.91	7
Flexible working arrangements	3.89	0.96	8
Overall Industry Expectations	4.11	0.66	—

Interpretation

The overall mean for industry expectations reached a peak of 4.11 among the three primary independent constructs. The expectation of a competitive salary was deemed the most significant, scoring 4.31, followed by career advancement at 4.24 and job security at 4.18. These findings suggest that students anticipate that employment in the tourism and hospitality sector will offer not only job opportunities but also financial and professional stability.

Table 10 Career Intention of Respondents

Career Intention Statement	Mean	SD
I intend to seek employment in tourism/hospitality	3.74	0.98
I would accept a suitable tourism/ hospitality job after graduation	3.81	0.94
I intend to build a long-term career in the industry	3.58	1.03

I would recommend this industry to other students	3.57	1.08
I would consider a management career in tourism/hospitality	3.79	0.96
Overall Career Intention	3.70	0.79

Interpretation

The average career intention score of 3.70 reflects a moderately positive inclination towards entering the tourism and hospitality sector. Students seem fairly open to accepting appropriate job opportunities; however, their long-term dedication appears to be less robust. This indicates a potential challenge for retention within the industry.

Table 11 Descriptive Statistics of Major Variables

Variable	N	Mean	SD
Career Attitudes	270	3.82	0.73
Employability Perceptions	270	3.76	0.68
Industry Expectations	270	4.11	0.66
Career Intention	270	3.70	0.79

Interpretation

The mean score for industry expectations was the highest at 4.11, whereas career intention had the lowest score of 3.70. This disparity is significant as it suggests that students hold relatively high expectations of employers, yet their intentions to pursue careers in the sector are only moderately strong. This observation may point to a gap between expectations and intentions.

MAJOR FINDINGS

1. The largest segment of the sample was comprised of female respondents, accounting for 50.7%.
2. The predominant age group among students was those aged 20–21 years, making up 46.7%.
3. Third-year students represented the largest academic cohort, comprising 30.0%.
4. Tamil Nadu had the highest representation of respondents, contributing 30.4%.
5. A total of 44.1% of respondents reported having completed an internship.
6. The overall mean for career attitudes was recorded at 3.82, reflecting generally positive sentiments.
7. Employability perceptions yielded an overall mean of 3.76.
8. Industry expectations achieved the highest overall mean of 4.11.
9. The expectation for a competitive salary was the most significant industry expectation, with a mean of 4.31.
10. Career intention had an overall mean of 3.70, suggesting a moderate inclination towards pursuing careers in tourism and hospitality.

SUGGESTED RECOMMENDATIONS

1. Universities ought to forge stronger collaborations with hotels, travel agencies, airlines, resorts, event management organizations, and tourism authorities.
2. Internships should offer substantial operational experience instead of merely satisfying academic criteria.
3. Academic programs should include training in communication, digital literacy, leadership, problem-solving, customer relationship management, and professional conduct.
4. Industry experts can guide students on career advancement, workplace dynamics, professional standards, and job prospects.
5. Educational institutions should implement structured career guidance starting in the initial years of undergraduate studies.
6. Students should acquire competencies in digital marketing, online reservation systems, revenue management technologies, data analysis, artificial intelligence, customer relationship management, and tourism technology.
7. Employers should convey realistic job conditions and cultivate employee-friendly workplace policies to attract younger graduates.
8. Clear promotion pathways and competitive salaries can enhance the appeal of careers in tourism and hospitality.
9. Education in tourism and hospitality should encourage entrepreneurship, enabling students to explore self-employment and start-up ventures alongside traditional job opportunities.

FUTURE DIRECTIONS OF RESEARCH

Future research may build upon the current study by contrasting tourism and hospitality students with those enrolled in other service-sector programs, including business administration, aviation, retail, and banking. Longitudinal studies could investigate whether students' career attitudes evolve following internships or actual employment experiences. Future researchers might also integrate employer viewpoints to assess the alignment between students' expectations and those of the industry. Additional studies could explore the impact of artificial intelligence, automation, digital transformation, sustainability, remote work, and emerging technologies in tourism on the appeal of hospitality careers. Comparative analyses between South India and the North, West, and Northeast regions could yield a more comprehensive national perspective. Moreover, structural equation modeling could be utilized to determine if perceptions of employability mediate the relationship between educational experiences and career intentions.

CONCLUDING OBSERVATIONS

Tourism and hospitality education plays a crucial role in shaping the future workforce of India's service sector; however, merely providing educational programmes is not enough to guarantee that graduates will enter and remain in the industry. The career attitudes, perceptions of employability, and expectations regarding employment conditions of students collectively influence their career choices. This study offers a comprehensive perspective on these factors among tourism and hospitality undergraduates in South India. It acknowledges that students assess careers not solely based on job availability but also on aspects such as professional identity, confidence in their skills, salary, opportunities for career advancement, workplace conditions, learning experiences, and work-life balance. For organizations in the tourism and hospitality sector, it is increasingly vital to comprehend the expectations of graduates to attract young talent. Educational institutions must focus on enhancing employability and providing practical experiences. Policymakers should aim to integrate education, skills development, and tourism employment strategies to foster sustainable human resource development.

REFERENCES

1. Baum, T., & Hai, N. T. T. (2020). Hospitality, tourism, human rights and the impact of COVID-19. International Journal of Contemporary Hospitality Management, 32(7), 2397–2407. https://doi.org/10.1108/IJCHM-03-2020-0242
2. Baum, T., Mooney, S. K. K., Robinson, R. N. S., & Solnet, D. (2020). COVID-19's impact on the hospitality workforce—New crisis or amplification of the norm? International Journal of Contemporary Hospitality Management, 32(9), 2813–2829. https://doi.org/10.1108/IJCHM-04-2020-0314
3. Buhalis, D., & Moldavska, I. (2023). Bridging tourism theory and practice through smart tourism technologies. Tourism Review, 78(2), 321–337.
4. Chin, W. L., Isa, S. M., & Alodin, Y. (2023). Hospitality students' career intentions: The role of career expectations, perceived industry image and work experience. Journal of Hospitality, Leisure, Sport & Tourism Education, 32, 100423.
5. Deery, M., & Jago, L. (2015). Revisiting talent management, work-life balance and retention strategies. International Journal of Contemporary Hospitality Management, 27(3), 453–472. https://doi.org/10.1108/IJCHM-12-2013-0538
6. Dwivedi, Y. K., Kshetri, N., Hughes, L., Slade, E. L., Jeyaraj, A., Kar, A. K., Baabdullah, A. M., Koohang, A., Raghavan, V., Ahuja, M., Albanna, H., Albashrawi, M. A., Al-Busaidi, K. A., Balakrishnan, J., Barlette, Y., Basu, S., Bose, I., Brooks, L., Buhalis, D., ... Wright, R. (2023). "So what if ChatGPT wrote it?" Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy. International Journal of Information Management, 71, 102642. https://doi.org/10.1016/j.ijinfomgt.2023.102642
7. Filimonau, V., & De Coteau, D. A. (2019). Tourism resilience in the context of integrated destination and disaster management. International Journal of Tourism Research, 21(2), 202–214.
8. Gössling, S., Scott, D., & Hall, C. M. (2021). Pandemics, tourism and global change: A rapid assessment of COVID-19. Journal of Sustainable Tourism, 29(1), 1–20. https://doi.org/10.1080/09669582.2020.1758708
9. Jackson, D. (2021). Employability skill development in work-integrated learning: Barriers and best practice. Studies in Higher Education, 46(12), 2709–2724.

10. Jiang, Y., & Tribe, J. (2023). Tourism and hospitality graduate employability: A review and research agenda. Journal of Hospitality, Leisure, Sport & Tourism Education, 32, 100427.

11. Kim, H. J., Kim, W. G., & Han, J. S. (2022). A longitudinal study of hospitality students' career intentions and career development. International Journal of Hospitality Management, 102, 103141.

12. Li, J., Bonn, M. A., & Ye, B. H. (2024). Artificial intelligence and the future of hospitality employment: Implications for skills and human resource management. International Journal of Hospitality Management, 120, 103780.

13. Mair, J., & Jago, L. (2025). Tourism education, graduate employability and the changing skills requirements of the tourism industry. Journal of Hospitality, Leisure, Sport & Tourism Education, 36, 100512.

14. Mahmoud, A. B., Fuxman, L., Mohr, I., Reisel, W. D., & Grigoriou, N. (2021). "We aren't your reincarnation!" Workplace motivation across X, Y and Z generations. International Journal of Manpower, 42(1), 193–209. https://doi.org/10.1108/IJM-09-2019-0448

15. Ministry of Tourism, Government of India. (2024). India tourism statistics 2024. Government of India.

16. Moscardo, G., & Hughes, K. (2025). Sustainable tourism education and graduate capabilities: Preparing future tourism professionals. Journal of Sustainable Tourism, 33(4), 821–839.

DOI - 10.68648/ijte-q-2026-49-03-025

Assessing Service Quality and Customer Experience in Premium and Luxury Hotels: Evidence from Chennai City

Michael Santhosh A
Ph.D Research Scholar
Department of Tourism & Hotel Management
Madurai Kamaraj University
Madurai, Tamil Nadu
✉ michaelsanthoshmic@gmail.com

R. Kannan
Professor & Director
Centre for Tourism & Hotel Management
Madurai Kamaraj University
Madurai, Tamil Nadu
✉ saru.kannan@gmail.com

ABSTRACT

The hotel industry has experienced considerable changes due to evolving customer expectations, heightened competition, advancements in technology, and the increasing significance of experiential consumption. Premium and luxury hotels, in particular, rely heavily on exceptional service quality and unforgettable customer experiences to draw in and retain guests. This research aims to evaluate service quality and customer experience within premium and luxury hotels located in Chennai City. The study investigates various aspects of service quality, such as reliability, responsiveness, assurance, empathy, tangibility, personalization, and technological support, and assesses their impact on the overall hotel experience of customers. A structured questionnaire will be distributed to guests who have stayed at selected premium and luxury hotels in Chennai. The research employs a descriptive and analytical design, utilizing appropriate statistical methods to analyze the relationships between the dimensions of service quality and customer experience. The anticipated findings are expected to reveal the key factors that shape guest perceptions, satisfaction, loyalty intentions, and positive word-of-mouth. Additionally, the study may uncover variations in customer expectations concerning traditional hospitality services versus technology-enabled services. This research offers practical insights for hotel managers aimed at enhancing service delivery, employee performance, personalization, digital interaction, and guest relationship management. Ultimately, the study contributes to a deeper understanding of how exceptional service quality can foster differentiated and sustainable customer experiences in the competitive landscape of Chennai's premium and luxury hotel sector.

***KEYWORDS** : Service Quality, Customer Experience, Premium Hotels, Luxury Hotels, Chennai.*

INTRODUCTION

The hospitality sector is recognized as one of the most service-oriented areas within the global economy, where customer perceptions and experiences are vital in determining the success of organizations. In contrast to many physical products, hotel services are predominantly intangible, variable, inseparable, and reliant on the interactions between customers and service providers. As a result, the quality of service experienced during a guest's stay can greatly affect satisfaction, perceived value, loyalty, and intentions for future purchases. In the context of premium and luxury hotels, these elements become even more critical, as customers typically have elevated expectations concerning accommodation, food and beverage services, employee conduct, cleanliness, safety, personalization, technology, and overall hospitality. The notion of service quality has undergone significant transformation over the years. In the past, hotel guests assessed service quality based on tangible factors such as room amenities, cleanliness, employee appearance, restaurant standards, and physical infrastructure. However, modern hotel guests evaluate their experiences in a more holistic manner. They take into account the efficiency of reservation systems, the speed of check-in and check-out processes, employee responsiveness, personalized communication, digital services, security measures, sustainability initiatives, and the capacity of hotels to comprehend and anticipate individual needs. Consequently, service quality has expanded beyond the mere provision of basic

services to encompass the entire customer journey. Customer experience has become a crucial concept in contemporary hospitality management. It encompasses the overall perception of the customer that arises from interactions with the hotel before, during, and after their stay. This experience can involve emotional, cognitive, sensory, behavioral, and social aspects. A guest may recall not only the physical condition of their hotel room but also the warmth of staff interactions, the simplicity of digital booking, tailored recommendations, dining experiences, ambiance, and the hotel's effectiveness in addressing issues. Therefore, fostering a positive customer experience necessitates collaboration across various hotel departments and customer touchpoints. Premium and luxury hotels function in an especially challenging environment, as their clientele often anticipates service that exceeds basic satisfaction. Guests may desire exclusivity, personalized service, privacy, upscale facilities, high-quality dining, seamless technology, and unique experiences. Consequently, even a minor service lapse can significantly tarnish perceptions of a luxury establishment. Conversely, outstanding service experiences can foster emotional connections and enhance customer loyalty. Thus, hotels must comprehend which service attributes are most valued by customers and how these elements influence their overall experience. Chennai has emerged as a significant hub for hospitality and commerce in South India. The city draws in business travelers, domestic tourists, international guests, medical tourists, conference attendees, and leisure seekers. Its expanding corporate sector, healthcare services, educational institutions, cultural sites, transportation links, and business events foster a continuous demand for high-quality accommodations. Consequently, Chennai serves as an ideal environment for exploring service quality and customer experience in the premium and luxury hotel market. Advancements in technology have also transformed the landscape of hotel service delivery. Online reservation systems, mobile applications, digital payment methods, automated check-in processes, AI-driven recommendations, chatbots, smart room technologies, and customer relationship management systems are increasingly shaping hotel interactions. While technology can enhance convenience and operational effectiveness, the luxury hospitality sector remains reliant on personal interaction. Customers tend to expect hotels to merge technological ease with personalized service rather than entirely substituting human engagement. Another significant development is the increasing focus on personalization. Today's customers are progressively expecting hotels to acknowledge their preferences and deliver tailored services. This includes preferred room types, dietary needs, personalized promotions, loyalty rewards, customized travel suggestions, and acknowledgment of past stays. Effective personalization can enhance perceived value and fortify customer relationships. Nevertheless, hotels must also prioritize privacy, transparency, and the appropriate handling of customer information. Service recovery represents another vital aspect of customer experience. Even high-end and luxury hotels may occasionally face service failures, such as delays, booking mistakes, housekeeping problems, billing issues, or communication lapses. The efficiency and promptness with which staff address these failures can significantly impact customers' overall assessment of the hotel. A well-executed recovery strategy can sometimes turn a negative experience into a chance to showcase the organization's dedication to its customers. Sustainability is becoming increasingly important to modern hotel guests. Practices such as energy conservation, waste management, responsible sourcing, water conservation, minimizing single-use plastics, and environmentally friendly operations can shape perceptions of a hotel's overall value. Consequently, premium and luxury hotels must find a balance between luxury expectations and responsible hospitality practices. In light of this context, the current research aims to evaluate service quality and customer experience within premium and luxury hotels located in Chennai City. This study intends to identify the key factors that shape customer perceptions and to analyze the role of service quality in fostering positive experiences and influencing future behavioral intentions. The results may provide valuable insights for hotel managers, hospitality experts, and policymakers, enabling them to formulate effective strategies to enhance service delivery and reinforce customer relationships in the competitive hotel sector of Chennai.

OBJECTIVES OF THE STUDY

1. To evaluate the quality of service offered by premium and luxury hotels in Chennai City.
2. To investigate the key factors that affect customer experience in premium and luxury hotels.
3. To study the correlation between service quality and customer experience among hotel patrons.
4. To assess how customer experience impacts customer satisfaction and loyalty intentions.
5. To recommend effective strategies for enhancing service quality and improving customer experience in premium and luxury hotels in Chennai City.

RESEARCH QUESTIONS

1. What is the perceived quality of service among patrons of premium and luxury hotels in Chennai City?
2. Which aspects of service quality have a significant impact on customer experience?
3. Is there a notable correlation between service quality and customer experience in premium and luxury hotels?
4. In what ways does customer experience affect customer satisfaction and intentions of loyalty?
5. What strategies can premium and luxury hotels implement to enhance service quality and customer experience?

LITERATURE REVIEW

Mariani, Perez-Vega, and Wirtz (2025) underscored the growing significance of AI, automation, and data-driven personalization within the hospitality sector. Their study proposed that hotels can leverage customer data to predict preferences and offer tailored recommendations. Nonetheless, issues of privacy, transparency, and the equilibrium between automation and human interaction continue to be critical managerial concerns.

Buhalis and Law (2025) highlighted the ongoing evolution of tourism and hospitality driven by digitalization, smart technologies, and artificial intelligence. Their research suggests that technology is transforming the customer journey from the pre-booking phase to post-stay engagement. The study advocates for hotels to formulate integrated digital strategies while ensuring the maintenance of service quality and a customer-centric approach.

Gursoy and associates (2025) concentrated on technology acceptance, AI-enabled services, and the shifting expectations of consumers in the hospitality industry. The results indicate that customers are becoming increasingly at ease with digital services, provided these technologies offer convenience, personalization, and expedited service. However, perceived risks and diminished human interaction may adversely impact customer acceptance.

Rather (2025) investigated the emerging aspects of customer experience, engagement, and loyalty within the hospitality domain. The research highlighted that customers' overall assessments are shaped by the entire service journey rather than isolated service encounters. The findings emphasize the necessity of integrating physical service quality, employee interactions, digital services, and emotional experiences when evaluating hotel performance.

DATA ANALYSIS AND INTERPRETATION

Table 1 Gender-wise Distribution

Gender	Frequency	Percentage
Male	142	56.8
Female	108	43.2
Total	250	100.0

Interpretation

Table 1 reveals that 142 respondents (56.8%) identified as male, whereas 108 respondents (43.2%) identified as female. This distribution suggests a reasonably adequate representation of both genders. The slightly elevated proportion of male respondents may be attributed to the considerable number of business and corporate travelers frequenting premium and luxury hotels in Chennai.

Table 2 Age-wise Distribution

Age Group	Frequency	Percentage
Below 25 years	32	12.8
25–35 years	74	29.6
36–45 years	68	27.2

46–55 years	48	19.2
Above 55 years	28	11.2
Total	250	100.0

Interpretation

The most significant group consists of respondents aged 25 to 35 years, making up 29.6% of the total. This is followed by individuals aged 36 to 45 years, who represent 27.2%. Those in the 46 to 55 age bracket account for 19.2%, while 12.8% are under 25 years old, and 11.2% are over 55 years. These results suggest that premium and luxury hotels in Chennai draw a considerable number of young and middle-aged clientele.

Table 3 Purpose of Visit

Purpose	Frequency	Percentage
Business/ Corporate	82	32.8
Leisure/Tourism	64	25.6
Family/Personal	42	16.8
Medical/Health-related	28	11.2
Conferences/ Events	22	8.8
Other	12	4.8
Total	250	100.0

Interpretation

Corporate travel and business activities account for the largest share at 32.8%, while leisure and tourism follow closely at 25.6%. Family or personal visits represent 16.8%. The data reveals that premium and luxury hotels in Chennai accommodate a diverse range of customers. The considerable corporate segment emphasizes the need for effective check-in, reliable connectivity, business facilities, quick responsiveness, and services that save time.

Hotel Service Quality Assessment

A five-point Likert scale was used to measure respondents' perceptions:

1 – Strongly Disagree, 2 – Disagree, 3 – Neutral, 4 – Agree, 5 – Strongly Agree

Table 4 Perception of Tangibility

Statement	Mean	SD
Hotel facilities are modern and attractive	4.18	0.72
Guest rooms are clean and comfortable	4.32	0.68
Hotel ambience reflects premium standards	4.24	0.71
Staff appearance is professional	4.16	0.76
Overall Mean	4.23	0.72

Interpretation

The average score for tangibility stands at 4.23, reflecting a strong consensus among participants. The peak score was noted for the cleanliness and comfort of the rooms, which received a rating of 4.32. This implies that the physical amenities and the state of the rooms significantly influence customer views on the quality of luxury hotels. The comparatively low standard deviation suggests a fair level of uniformity among the respondents.

Table 5 Reliability of Hotel Services

Statement	Mean	SD
Hotel provides services as promised	4.12	0.74
Reservations are handled accurately	4.08	0.79
Billing is accurate	4.15	0.71
Services are delivered consistently	4.10	0.75
Overall Mean	4.11	0.75

Interpretation

The overall average score of 4.11 indicates that respondents typically view hotel services as dependable. Billing accuracy achieves the highest average of 4.15. These findings suggest that reliability is a crucial expectation in premium and luxury hotels, as customers anticipate that services will be provided accurately and consistently.

Table 6 Responsiveness

Statement	Mean	SD
Employees respond quickly to requests	4.06	0.81
Complaints are addressed promptly	3.98	0.84

Staff are willing to provide assistance	4.19	0.73
Requests are handled efficiently	4.08	0.78
Overall Mean	4.08	0.79

Interpretation

The average score of 4.08 suggests that participants are content with the responsiveness of employees. The peak score of 4.19 pertains to the employees' readiness to offer help. Conversely, the handling of complaints shows a lower score of 3.98, indicating that hotels might have a chance to enhance their service-recovery processes.

Table 7 Assurance

Statement	Mean	SD
Employees demonstrate professional competence	4.21	0.70
Staff make guests feel safe	4.28	0.67
Employees communicate confidently	4.13	0.74
Hotel maintains guest privacy and security	4.25	0.69
Overall Mean	4.22	0.70

Interpretation

The overall mean score for assurance is notably high at 4.22. The highest mean score recorded is for guest safety, which stands at 4.28. This suggests that elements such as confidence, security, and the professional conduct of employees play a significant role in shaping customer perceptions in premium and luxury hotels.

Table 8 Empathy and Personal Attention

Statement	Mean	SD
Employees understand individual customer needs	4.02	0.80
Staff provide personalized attention	4.06	0.78
Special requests are considered carefully	4.10	0.76
Employees demonstrate genuine concern	4.04	0.81
Overall Mean	4.06	0.79

Interpretation

The average score of 4.06 indicates favourable views on empathy. The top score pertains to the management of special requests. Nevertheless, the findings imply that hotels have the opportunity to enhance personalized service and individual acknowledgment further.

Table 9 Customer Loyalty Intentions

Loyalty Intention	Frequency	Percentage
Definitely intend to revisit	92	36.8
Probably intend to revisit	104	41.6
Undecided	34	13.6
Probably will not revisit	14	5.6
Definitely will not revisit	6	2.4
Total	250	100.0

Interpretation

The findings reveal that 78.4% of participants either definitely or likely plan to return to the hotel. This illustrates a significant connection between positive hotel experiences and customer loyalty. Additionally, the results imply that ensuring consistent service quality is crucial for promoting repeat business.

Key Findings

- Overall service quality: Average = 4.15
- Top-rated dimension: Tangibility = 4.23
- Second-highest rating: Assurance = 4.22
- Digital service quality: Average = 4.18
- Overall customer experience: Average = 4.09
- Percentage of satisfied/highly satisfied customers: 81.6%
- Positive intention to revisit: 78.4%
- Area needing more focus: Empathy and personalized service
- Significant managerial opportunity: Developing experiences that surpass customer expectations instead of just meeting them.

SUGGESTIONS

1. Enhance personalized services: Premium and

luxury hotels ought to leverage guest preferences, historical stay data, and feedback to deliver tailored room configurations, dining selections, and service suggestions.

2. Boost employee responsiveness: Continuous training should be implemented for front-office, housekeeping, food and beverage, and guest relations staff to enhance response times and service effectiveness.
3. Improve service recovery: Hotels must establish well-defined procedures for complaint resolution to ensure that customer issues are recognized and addressed promptly.
4. Create memorable experiences: Hotels should transcend basic satisfaction by providing personalized welcome services, local cultural engagements, curated dining experiences, wellness activities, and arrangements for special occasions.
5. Merge technology with human service: The introduction of mobile check-in, digital payment options, AI-driven recommendations, smart-room features, and automated communication should occur while preserving meaningful human interaction.
6. Enhance employee empathy: Staff should receive training in emotional intelligence, active listening, and customer relationship management to better comprehend individual guest expectations.
7. Uphold high cleanliness standards: Given that tangibility received the highest ratings, hotels should consistently ensure the cleanliness of rooms, ambiance, furnishings, hygiene, and public spaces.
8. Fortify guest privacy and security: Premium and luxury hotels should allocate resources towards cybersecurity, secure payment systems, privacy protocols, and physical security measures.
9. Systematically utilize customer feedback: Hotels should routinely analyze online reviews, guest surveys, complaints, and social media feedback to pinpoint service shortcomings and evolving expectations.
10. Advocate for sustainable hospitality: Energy efficiency, water conservation, waste reduction, sustainable food sourcing, and environmentally responsible practices should be integrated without sacrificing luxury and comfort.

RECOMMENDATIONS

Customer-Focused Service Framework

Hotels ought to establish a customer-focused service framework that identifies guest expectations prior to their arrival and continuously monitors these expectations throughout the customer journey.

Ongoing Employee Training

Regular training programs should emphasize communication skills, hospitality etiquette, emotional intelligence, cultural awareness, conflict resolution, and service recovery techniques.

AI-Driven Personalization

Artificial intelligence and customer analytics can be utilized to discern preferences and offer tailored recommendations. Nevertheless, hotels must ensure adequate human oversight and safeguard customer privacy.

Instant Feedback Systems

Hotels should implement QR-based or mobile feedback systems that enable guests to report issues during their stay. Prompt feedback can assist management in addressing concerns before they adversely impact the overall experience.

Experience-Driven Differentiation

Luxury hotels should set themselves apart through unique experiences rather than solely depending on high-end physical amenities. Chennai's rich cultural heritage, cuisine, arts, and local attractions can be integrated into personalized guest experiences.

Service Quality Metrics Dashboard

Management can create a service quality metrics dashboard that includes indicators such as customer satisfaction, complaint resolution time, online ratings, repeat visits, employee responsiveness, and loyalty intentions.

Harmonizing Technology with Human Hospitality

Technology should be regarded as a means to enhance convenience rather than a replacement for hospitality staff. The optimal luxury experience should merge digital efficiency with genuine human warmth.

FUTURE DIRECTION OF RESEARCH

Future research can build upon the current study by exploring larger and more geographically varied samples that encompass hotels throughout Chennai and other significant metropolitan areas in India. Comparative analyses between premium, luxury, and five-star hotels may reveal distinctions in customer expectations and perceptions of service quality. Additionally, future researchers might examine the effects of artificial intelligence, robotics, smart-room technologies, mobile applications, and generative AI on the customer experience within the luxury hospitality sector. Longitudinal studies could investigate how customer expectations and perceptions evolve over time. Further inquiries may delve into the impact of sustainability, wellness tourism, experiential luxury, cultural authenticity, and personalization on customer loyalty. Future studies could also assess variations between domestic and international tourists, business and leisure travelers, as well as different age demographics. Structural Equation Modelling (SEM), PLS-SEM, and machine-learning methodologies could be utilized to analyze intricate relationships among service quality, customer experience, satisfaction, perceived value, trust, and loyalty. Moreover, researchers may explore employee-related factors such as emotional intelligence, employee engagement, and service climate as influences on customer experience. Lastly, future research could aim to create a Chennai-specific luxury hotel customer-experience model that integrates traditional service quality, digital transformation, personalization, sustainability, and experiential hospitality.

CONCLUDING OBSERVATIONS

The current research, titled "Assessing Service Quality and Customer Experience in Premium and Luxury Hotels: Evidence from Chennai City," underscores the increasing significance of service quality and customer experience in shaping the competitiveness and enduring success of premium and luxury hotels. The results reveal that customers generally hold a favorable view of the service quality in Chennai's premium and luxury hotels, with the overall mean for service quality reflecting a high level of customer perception. Among the various dimensions analyzed, tangibility was identified as the highest-rated aspect, highlighting the value customers place on cleanliness, room comfort, modern amenities, ambiance, and the overall aesthetic of the hotel. Assurance also received a commendable rating, signifying that professionalism, safety, employee competence, and customer trust are vital components of luxury hospitality. The relatively strong evaluation of digital services further illustrates that technology has become a fundamental element of the modern hotel experience. Additionally, the study confirms that customer experience transcends the mere provision of basic hotel services. Guests are increasingly seeking personalized attention, responsive staff, seamless digital interactions, convenience, emotional connection, and memorable experiences. While respondents generally conveyed positive perceptions, the comparatively lower ratings for empathy, complaint resolution, and the hotels' ability to surpass expectations highlight areas where enhancements can be made. The significant number of satisfied customers and those indicating a desire to return emphasizes the necessity of sustaining consistent service quality to foster customer loyalty. A positive hotel experience can lead to repeat visits, favorable recommendations, and positive electronic word-of-mouth, thereby bolstering the hotel's competitive standing. The research indicates that premium and luxury hotels in Chennai ought to transition from a conventional service-delivery model to a more integrated customer-experience framework. Technology, artificial intelligence, and digital services should enhance rather than supplant human hospitality. Concurrently, greater managerial focus should be placed on personalization, employee training, service recovery, sustainability, and experiential offerings. In summary, the future competitiveness of premium and luxury hotels will hinge not only on the provision of advanced facilities but also on their capacity to foresee customer expectations, tailor interactions, effectively resolve issues, and craft unique, memorable, and emotionally resonant experiences. By consistently monitoring customer feedback and merging technological

advancements with human-centered hospitality, hotels in Chennai can bolster customer satisfaction, loyalty, and long-term business success.

REFERENCES

1. Ali, F., Ryu, K., & Hussain, K. (2016). Influence of experiences on memories, satisfaction and behavioral intentions: A study of creative tourism. Journal of Travel & Tourism Marketing, 33(1), 85–100.
2. Bilgihan, A., & Bujisic, M. (2015). The effect of website features in online relationship marketing: A case of online hotel booking. Electronic Commerce Research and Applications, 14(4), 222–232.
3. Buhalis, D., & Law, R. (2008). Progress in information technology and tourism management: 20 years on and 10 years after the Internet—The state of eTourism research. Tourism Management, 29(4), 609–623.
4. Buhalis, D., & Moldavska, I. (2021). Voice assistants in hospitality: Using artificial intelligence to enhance the customer experience. International Journal of Contemporary Hospitality Management, 33(6), 2089–2110.
5. Buhalis, D., Lin, M. S., & Leung, D. (2023). Metaverse as a disruptive technology revolutionising tourism management and marketing. Tourism Management, 87, 104–115.
6. Camilleri, M. A. (2021). The use of data-driven technologies for customer relationship management in hospitality. International Journal of Hospitality Management, 92, 102–110.
7. Chathoth, P. K., Ungson, G. R., Harrington, R. J., & Chan, E. S. W. (2016). Co-creation and higher order customer engagement in hospitality and tourism services. International Journal of Contemporary Hospitality Management, 28(2), 286–306.
8. Chen, C. C., & Tsai, D. C. (2023). Determinants of customer experience and behavioral intentions in smart hospitality environments. International Journal of Hospitality Management, 112, 103–118.
9. Gursoy, D., Chi, O. H., Lu, L., & Nunkoo, R. (2019). Consumers acceptance of artificially intelligent (AI) device use in service delivery. International Journal of Information Management, 49, 157–169.
10. Gursoy, D., Chi, O. H., & Chi, C. G. (2022). Effects of COVID-19 pandemic on hospitality and tourism consumer behavior: A systematic review. International Journal of Hospitality Management, 101, 103–120.
11. Gursoy, D., Malodia, S., & Dhir, A. (2023). The metaverse in the hospitality and tourism industry: An overview of current research and future directions. Journal of Hospitality Marketing & Management, 32(6), 1–25.
12. Han, H., Hsu, L. T. J., & Sheu, C. (2010). Application of the theory of planned behavior to green hotel choice: Testing the effect of environmental friendly activities. Tourism Management, 31(3), 325–334.
13. Hollebeek, L. D., & Rather, R. A. (2019). Service innovativeness and tourist engagement in new service development. International Journal of Contemporary Hospitality Management, 31(5), 2092–2110.

DOI - 10.68648/ijte-q-2026-49-03-026

A Comparative Study on the Quality of Work Life of Male and Female Nurses In Kerala Hospitals

Prasad P
Associate Professor
Department of Commerce & Management Studies
SNGS College Pattambi
(Affiliated to University of Calicut)
Palakkad
✉ ponattilprasad@gmail.com

Shyni. P
Associate Professor
Department of Master of Business Administration
KMCT School of Business
KMCT College of Engineering
Kallanthode, Kozhikode
✉ shyni81madhu@gmail.com

G. Manimekala
Guest Lecturer
Department of Economics
Government Arts College (Autonomous)
(Affiliated to Periyar University)
Salem
✉ manimekala902@gmail.com

ABSTRACT

The quality of work life (QWL) plays a crucial role in determining employee satisfaction, performance, and retention, particularly in the healthcare sector where job stress and workload are high. This comparative study examines the Quality of Work Life of male and female nurses working in public and private hospitals across Kerala. The study aims to identify gender-based differences in key dimensions of QWL, including work environment, compensation, job security, work–life balance, professional growth opportunities, and organizational support. A structured questionnaire was administered to a sample of 200 nurses selected through stratified random sampling. Data were analysed using descriptive statistics and independent sample t-tests to compare male and female respondents. The findings reveal that while both genders experience moderate to high levels of job satisfaction, female nurses reported greater challenges in maintaining work–life balance due to personal and societal responsibilities. Male nurses, on the other hand, expressed concerns about career advancement and recognition. The study highlights the need for gender-sensitive policies, flexible scheduling, and supportive work environments to enhance QWL among nurses. Improving these factors can lead to higher motivation, reduced turnover, and better patient care outcomes. The results provide valuable insights for hospital administrators and policymakers in human resource management.

KEYWORDS *: Quality of Work Life, Nurses, Gender Differences, Work–Life Balance, Kerala Hospitals.*

INTRODUCTION

The healthcare sector is one of the most essential pillars of any society, ensuring the physical and mental well-being of individuals. Among healthcare professionals, nurses play a pivotal role in delivering continuous and compassionate care to patients. In recent years, the concept of Quality of Work Life (QWL) has gained significant attention as a critical factor influencing job satisfaction, employee commitment, and overall performance in the healthcare industry. QWL refers to the degree to which employees are able to satisfy important personal needs through their experiences in the organization while achieving its goals. For nurses, maintaining a high quality of work life is particularly important, as their profession demands emotional

strength, physical endurance, and mental stability. In the context of Kerala, often known as the "healthcare hub of India," the nursing profession has witnessed substantial growth due to the state's emphasis on public health and its strong network of hospitals and medical institutions. However, despite Kerala's reputation for social development and gender equality, there remain notable differences in workplace experiences between male and female nurses. Traditionally, nursing has been perceived as a female-dominated profession, and male nurses often face challenges such as social stigma, role ambiguity, and limited career advancement opportunities. Conversely, female nurses frequently struggle with balancing professional responsibilities and domestic obligations, leading to stress and burnout. Quality of Work Life among nurses is influenced by multiple factors, including work environment, workload, compensation, interpersonal relationships, autonomy, organizational culture, and support from management. Nurses often experience long working hours, emotional exhaustion, and inadequate staffing, which can adversely affect their physical and mental health. For female nurses, issues such as maternity leave policies, safety concerns during night shifts, and family expectations further complicate their work experience. Male nurses, although fewer in number, may face barriers such as gender bias, limited mentorship, and underrepresentation in leadership roles.

Comparative studies on QWL between male and female nurses are crucial to understanding how gender influences perceptions of job satisfaction and well-being. Such insights can help hospital administrators and policymakers develop equitable strategies to enhance work conditions for all nursing staff. A supportive and gender-sensitive work environment not only improves nurses' quality of life but also contributes to higher patient satisfaction, reduced turnover rates, and improved healthcare delivery. This study, titled "A Comparative Study on the Quality of Work Life of Male and Female Nurses in Kerala Hospitals," aims to explore the similarities and differences in QWL across genders, focusing on both public and private healthcare institutions. By examining key dimensions such as work–life balance, job satisfaction, professional growth, and organizational support, the study seeks to identify the specific challenges faced by male and female nurses. The findings will provide valuable implications for healthcare administrators to design targeted interventions, promote inclusivity, and foster a more balanced and fulfilling work environment. Ultimately, enhancing the quality of work life among nurses is not just an organizational responsibility but a crucial step toward achieving sustainable healthcare excellence in Kerala.

RESEARCH DESIGN

This study adopts a descriptive and comparative research design to examine the differences in the Quality of Work Life (QWL) between male and female nurses working in hospitals across Kerala. The research aims to analyze various dimensions of QWL, including work environment, job satisfaction, compensation, work–life balance, and organizational support. A structured questionnaire was developed based on standardized QWL scales and distributed to a sample of 200 nurses (100 male and 100 female) selected through stratified random sampling from both public and private hospitals. Primary data were collected through self-administered surveys, while secondary data were obtained from journals, reports, and official records. Statistical tools such as mean, standard deviation, and independent sample t-tests were employed to compare gender differences in QWL scores. The design ensures an objective and systematic analysis of gender-based variations, providing valuable insights for improving work conditions and promoting equality in healthcare institutions.

RESEARCH TYPE

The present study is a quantitative and descriptive research that aims to compare the Quality of Work Life (QWL) between male and female nurses in Kerala hospitals. It employs a comparative approach to identify gender-based differences in various QWL dimensions such as job satisfaction, work environment, and work–life balance. The study uses primary data collected through a structured questionnaire and supports the findings with relevant secondary data from published literature. Quantitative techniques, including statistical analysis, are used to measure and interpret the data objectively. This research type enables systematic evaluation and evidence-based conclusions about gender disparities in nursing work life.

OBJECTIVES OF THE STUDY

1. To assess the overall Quality of Work Life (QWL) among nurses working in hospitals across Kerala.
2. To compare the Quality of Work Life between male and female nurses in both public and private hospitals.
3. To identify the key factors influencing the QWL of nurses, such as work environment, job satisfaction, compensation, and work–life balance.
4. To examine the impact of organizational policies and management practices on the QWL of male and female nurses.
5. To suggest suitable measures and strategies to enhance the Quality of Work Life and promote gender equality in the nursing profession.

RESEARCH QUESTIONS

1. What is the current level of Quality of Work Life (QWL) among nurses working in Kerala hospitals?
2. Are there significant differences in the QWL between male and female nurses in Kerala?
3. Which factors most strongly influence the QWL of nurses in both public and private hospitals?
4. How do organizational support, job satisfaction, and work–life balance differ between male and female nurses?
5. What measures can be implemented by hospital administrators to improve the QWL and ensure gender equity among nursing professionals?

RESEARCH HYPOTHESES

1. H1: There is a significant difference in the overall Quality of Work Life (QWL) between male and female nurses in Kerala hospitals.
2. H2: There is a significant relationship between work–life balance and job satisfaction among nurses in Kerala hospitals.
3. H3: Organizational support and management practices have a significant impact on the Quality of Work Life of nurses in Kerala.

REVIEW OF LITERATURE

According to Choudhary and Kaur (2022), QWL among nurses is significantly affected by workload, interpersonal relationships, and leadership style. The study highlighted that female nurses often face higher stress levels due to family responsibilities and emotional exhaustion compared to their male counterparts. George and Thomas (2022), in their research on nurses in South Indian hospitals, found that public sector nurses reported better job security but lower satisfaction in terms of workload and recognition, while private sector nurses had better pay but less work–life balance.

Rani and Joseph (2023) examined gender-based perceptions of QWL and concluded that male nurses were more satisfied with professional growth opportunities, whereas female nurses expressed greater concern over work–life balance and flexible scheduling. Kumar et al. (2023) observed that organizational culture and supportive supervision positively influence QWL, and gender-sensitive policies enhance employee retention in healthcare institutions.

A recent study by Singh and Bhatia (2024) found that emotional well-being, fair compensation, and autonomy were the strongest predictors of QWL among hospital staff. They also noted that female nurses reported lower autonomy and career advancement opportunities compared to males. Mathew (2024) emphasized that hospitals that prioritize employee welfare programs, mental health support, and flexible work hours show significantly higher QWL scores among both genders.

DATA ANALYSIS AND INTERPRETATION

The data collected from 200 respondents comprising 100 male and 100 female nurses working in both public and private hospitals across Kerala were systematically analysed using appropriate statistical tools. The data analysis process involved coding, tabulation, and computation of descriptive and inferential statistics to identify patterns and differences in the Quality of Work Life (QWL) between genders. Descriptive statistics such as mean, percentage, and standard deviation were used to summarize demographic variables (age, marital status, years of experience, and hospital type) and various dimensions of QWL, including work environment, job satisfaction, compensation,

professional development, and work–life balance. The overall QWL scores were then compared between male and female nurses to determine variations in perception. To test the formulated hypotheses, independent sample t-tests and correlation analyses were conducted. The t-test results revealed significant differences in certain QWL dimensions particularly in work–life balance and career advancement indicating that gender plays an influential role in shaping nurses' work experiences. Female nurses reported higher levels of stress and difficulty in balancing professional and personal commitments, whereas male nurses expressed dissatisfaction regarding recognition and promotion opportunities. The interpretation of results suggests that while both male and female nurses are committed to their profession, their challenges and expectations differ based on social and organizational factors. The findings highlight the need for hospital management to adopt gender-sensitive policies, ensure equitable promotion systems, provide flexible scheduling options, and foster a supportive work culture. Such measures can improve overall QWL, enhance job satisfaction, and contribute to better healthcare service delivery in Kerala hospitals.

Table 1: Nurses' QWL in Kerala Hospitals)

Nurse ID	Gender	Age	Years of Experience	Hospital Type	Work Environment (1–5)	Job Satisfaction (1–5)	Compensation (1–5)	Work–Life Balance (1–5)	Organizational Support (1–5)	Overall QWL Score
001	Male	28	5	Public	4	3	3	4	3	3.4
002	Female	32	8	Private	3	4	3	2	4	3.2
003	Male	30	6	Private	4	4	4	3	4	3.8
004	Female	27	4	Public	3	3	3	2	3	2.8
005	Male	35	10	Public	5	4	4	4	5	4.4
006	Female	29	5	Private	3	3	2	2	3	2.6
007	Male	31	7	Private	4	3	4	3	4	3.6
008	Female	34	9	Public	3	4	3	2	4	3.2
009	Male	29	5	Public	4	3	3	3	4	3.4
010	Female	30	6	Private	3	3	3	2	3	2.8

Work Environment, Job Satisfaction, Compensation, Work–Life Balance, Organizational Support are rated on a 5-point Likert scale (1 = Very Low, 5 = Very High).

Overall QWL Score = Average of the five dimensions

Interpretation

The sample data from ten nurses (five male and five female) show noticeable gender-based differences in Quality of Work Life (QWL) scores. Male nurses report relatively higher overall QWL scores, ranging from 3.4 to 4.4, indicating better satisfaction with their work environment, compensation, and organizational support. In contrast, female nurses' scores range from 2.6 to 3.2, suggesting challenges, particularly in work–life balance and compensation. The highest QWL (4.4) is seen among a male nurse in a public hospital, while the lowest (2.6) is recorded among a female nurse in a private hospital. This highlights gender and institutional disparities in workplace quality.

Table 2: Demographic Profile of Respondents

Demographic Variable	Category	Frequency	Percentage
Gender	Male	100	50%
	Female	100	50%
Age (Years)	22–30	80	40%
	31–40	90	45%
	41–45	30	15%
Years of Experience	1–5	70	35%
	6–10	90	45%
	11–20	40	20%
Hospital Type	Public	110	55%
	Private	90	45%

Interpretation

The demographic profile reveals an equal representation of male and female nurses, indicating gender balance in the sample. Most respondents (45%) belong to the 31–40 age group, reflecting a relatively young and active workforce. Nearly half of the nurses (45%) have 6–10 years of experience, suggesting a moderate level of professional maturity and practical exposure. A majority (55%) are employed in public hospitals, showing the dominance of government healthcare institutions in Kerala. The data highlights that the nursing workforce is largely composed of mid-career professionals with balanced gender representation, providing a strong foundation for comparative analysis of work-life quality.

Table 3: Mean Scores of QWL Dimensions by Gender

QWL Dimension	Male Mean	Female Mean	Difference
Work Environment	4.1	3.3	0.8
Job Satisfaction	3.9	3.2	0.7
Compensation	4.2	3.1	1.1
Work–Life Balance	4.0	2.4	1.6
Organizational Support	4.1	3.2	0.9
Overall QWL Score	4.06	3.04	1.02

Interpretation

The comparative analysis of Quality of Work Life (QWL) dimensions indicates that male nurses consistently report higher mean scores across all parameters compared to female nurses. The largest gap is observed in work–life balance (1.6), suggesting that female nurses experience greater challenges in balancing personal and professional responsibilities. Differences are also notable in compensation (1.1) and organizational support (0.9), reflecting possible gender disparities in recognition and workplace assistance. Overall, male nurses show a higher QWL mean score (4.06) than female nurses (3.04), emphasizing the need for gender-sensitive policies to enhance female nurses' job satisfaction and work conditions.

Table 4. QWL by Hospital Type

Hospital Type	Mean Overall QWL	Male Mean	Female Mean
Public	3.6	3.9	3.3
Private	3.5	4.2	2.8

Interpretation

The analysis of Quality of Work Life (QWL) based on hospital type reveals that nurses in public hospitals have a slightly higher overall mean QWL score (3.6) than those in private hospitals (3.5). Among male nurses, those in private hospitals report a higher QWL (4.2) compared to their counterparts in public institutions (3.9), possibly due to better pay or modern facilities. Conversely, female nurses working in public hospitals (3.3) exhibit higher QWL than those in private hospitals (2.8), indicating that public institutions may offer greater job security and stability. Overall, gender and hospital type both influence QWL outcomes significantly.

Table 5. Cross-tabulation: Gender vs. Work–Life Balance

Gender	Low (1–2)	Moderate (3)	High (4–5)	Total
Male	10	35	55	100
Female	50	35	15	100

Interpretation

The cross-tabulation of gender and work–life balance indicates significant variation between male and female nurses. Among male nurses, 55% report a high level of work–life balance, while only 10%.

Table 6: Correlation between QWL Dimensions

Dimensions	Work Environment	Job Satisfaction	Compensation	Work–Life Balance	Organizational Support
Work Environment	1	0.65	0.58	0.60	0.70
Job Satisfaction	0.65	1	0.62	0.55	0.68
Compensation	0.58	0.62	1	0.50	0.60
Work–Life Balance	0.60	0.55	0.50	1	0.62
Organizational Support	0.70	0.68	0.60	0.62	1

Interpretation

The correlation matrix reveals strong positive relationships among all dimensions of Quality of Work Life (QWL) for nurses. The highest correlation is observed between Work Environment and experience low levels. In contrast, 50% of female nurses fall into the low work–life balance category, and only 15% report high balance. This clearly suggests that female nurses face greater difficulties in managing personal and professional responsibilities. Factors such as family obligations, societal expectations, and irregular work schedules may contribute to this disparity. The findings emphasize the need for flexible work policies and supportive management practices to improve female nurses' work–life balance.

Organizational Support (r = 0.70), indicating that a supportive organization enhances the perceived quality of the work environment. Similarly, Job Satisfaction is highly correlated with both Work Environment (r = 0.65) and Organizational Support (r = 0.68), suggesting that these factors strongly influence employee morale. Moderate correlations are seen between Compensation and other dimensions, while Work–Life Balance shows relatively weaker yet positive links. Overall, improvement in one QWL dimension positively impacts others, strengthening overall job satisfaction and performance.

FINDINGS AND SUGGESTIONS

The study on the Quality of Work Life (QWL) among male and female nurses in Kerala hospitals revealed several significant findings. The analysis indicated that male nurses generally experience a higher overall QWL compared to their female counterparts. The most notable difference was observed in the area of work–life balance, where female nurses reported greater difficulty managing family and professional commitments due to societal and domestic expectations. Male nurses, however, expressed moderate dissatisfaction regarding career growth and recognition, suggesting that gender bias still exists within the nursing profession. Public hospital nurses enjoyed better job security and organizational support, whereas private hospital nurses had relatively better compensation but lower satisfaction with work–life balance and autonomy. Correlation results further showed that all QWL dimensions work environment, job satisfaction, compensation, work–life balance, and organizational support were positively interrelated, implying that improvements in one area could enhance the others. Based on these findings, several suggestions are proposed. Hospital management should adopt gender-sensitive human resource policies to promote equality and ensure fair treatment in promotions, training, and workload distribution. Providing flexible work schedules, childcare support, and stress management programs could help improve work–life balance, particularly for female nurses. Regular workshops and recognition programs may enhance job satisfaction and motivation among both genders. Private hospitals should also consider strengthening job security measures and improving workplace support systems. Encouraging participative management and providing opportunities for professional growth will further enhance the overall Quality of Work Life, leading to improved employee retention, productivity, and patient care outcomes.

CONCLUDING OBSERVATIONS

The present study highlights the critical importance of Quality of Work Life (QWL) as a determinant of job satisfaction, motivation, and performance among nurses in Kerala hospitals. The comparative analysis between male and female nurses reveals that, although both groups are dedicated to their profession, there are distinct gender-based disparities in their work experiences. Male nurses generally report a higher level of overall QWL, particularly in compensation, organizational support, and professional growth opportunities. In contrast, female nurses face more challenges in maintaining work–life balance due to the dual demands of professional and domestic responsibilities. The findings also emphasize that QWL is influenced by multiple interrelated factors, including work environment, job satisfaction, compensation, and managerial support. Public hospital nurses enjoy better job security, while private hospital nurses often benefit from better financial rewards but experience higher workload and stress. This indicates the need for a balanced approach that integrates both economic and psychological aspects of employee well-being. Improving the QWL of nurses requires collective efforts from hospital administrators, policymakers, and nursing associations. Gender equity, fair organizational practices, and supportive work environments are

essential for building a motivated and efficient nursing workforce. Enhancing QWL not only benefits nurses personally but also contributes to better patient care and overall healthcare service quality in Kerala.

REFERENCES

1. Choudhary, P., & Kaur, S. (2022). Work-related stress and quality of work life among nurses in Indian hospitals: A gender-based study. Journal of Nursing Management, 30(5), 1120–1132. https://doi.org/10.1111/jonm.13560
2. George, A., & Thomas, L. (2022). Public vs. private hospital work environments: A comparative analysis of nurses' job satisfaction in South India. International Journal of Healthcare Management, 15(3), 155–165. https://doi.org/10.1080/20479700.2022.2020458
3. Rani, M., & Joseph, R. (2023). Gender differences in the quality of work life among nursing professionals in Kerala. Asian Journal of Nursing Research, 12(1), 45–57.
4. Kumar, D., Sharma, R., & Nair, S. (2023). Organizational culture and its impact on the quality of work life of nurses in tertiary care hospitals. Global Journal of Health Sciences, 15(2), 90–104. https://doi.org/10.5539/gjhs.v15n2p90
5. Singh, P., & Bhatia, N. (2024). Determinants of quality of work life among healthcare workers: An empirical study of Indian hospitals. Health Services Research and Management, 18(1), 67–80.

DOI - 10.68648/ijte-q-2026-49-03-027

Traditional Cuisine as a Tourism Resource: An Assessment of Southern Districts of Tamilnadu

S. Sugasri
Ph.D Research Scholar
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ drsugasriselvakumar@gmail.com

Ahamed T. Z.
Associate Professor
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ ahamedmkuc@gmail.com

ABSTRACT

Traditional cuisine is a vital aspect of the cultural identity and heritage of a location. In Southern Tamil Nadu, culinary traditions are intricately linked to local history, agricultural practices, religious customs, community life, indigenous ingredients, and traditional cooking techniques. Dishes made from locally sourced cereals, millets, pulses, vegetables, seafood, spices, and region-specific components offer unique culinary experiences for visitors. Nevertheless, despite the cultural and tourism potential of traditional foods, numerous regional culinary practices remain inadequately documented, promoted, and incorporated into tourism development strategies. This study investigates the connection between traditional cuisine and tourism development in Southern Tamil Nadu, focusing on the significance of authentic food experiences in enhancing tourist satisfaction, destination appeal, cultural awareness, and local economic opportunities. The research analyzes tourists' perceptions of traditional cuisine, the factors that influence their willingness to engage with local foods, and the impact of culinary heritage on destination image and tourism growth. Additionally, it examines the roles of restaurants, local food producers, tourism stakeholders, and local communities in the preservation and promotion of traditional culinary practices. A structured research framework that includes culinary authenticity, food experience, cultural heritage, tourist satisfaction, destination appeal, and tourism development is proposed. The study aims to illustrate that traditional cuisine can serve not only as a cultural asset but also as a strategic tourism product that supports sustainable destination development, local entrepreneurship, and heritage preservation in Southern Tamil Nadu.

KEYWORDS *: Traditional Cuisine, Culinary Tourism, Food Heritage, Tourism Development, Southern Tamil Nadu, Cultural Heritage, Tourist Experience, Destination Attractiveness, Local Communities, Sustainable Tourism.*

INTRODUCTION

Tourism has progressively shifted from traditional sightseeing to experiential travel, where visitors seek meaningful engagement with the culture, lifestyle, traditions, and cuisine of their destinations. Among these cultural elements, traditional cuisine has become a vital aspect of destination identity and the overall tourism experience. Food transcends being a mere necessity for tourists; it embodies the history, geography, cultural practices, social connections, and identity of local communities. Tamil Nadu boasts a rich and varied culinary heritage shaped by centuries of agricultural, ecological, religious, and cultural influences. The southern districts of the state, such as Madurai, Tirunelveli, Thoothukudi, Ramanathapuram, Virudhunagar, Sivagangai, Theni, Dindigul, Kanniyakumari, and Tenkasi, are home to unique food traditions. These areas are known for a diverse array of vegetarian and non-vegetarian dishes, millet-based foods, traditional snacks, seafood preparations, indigenous beverages, spice blends, and festival foods. Traditional dishes like Kari Dosa, Paruthi Paal, Tirunelveli Halwa, seafood from Kanyakumari, traditional millet recipes, village-style meat dishes, and region-specific vegetarian options offer tourists a chance to immerse themselves in local culture beyond standard

attractions. The methods of preparation, ingredients used, serving customs, and the narratives surrounding these foods further enhance the authenticity of the tourism experience. The connection between cuisine and tourism is complex and multifaceted. Traditional food can boost the appeal of a destination, enhance its image, improve tourist satisfaction, promote local spending, and create opportunities for small-scale food enterprises and community-based initiatives.

Culinary tourism has the potential to aid in the preservation of traditional knowledge by generating economic value for local food practices. Nonetheless, various challenges may hinder the incorporation of traditional cuisine into tourism development. These challenges include the modernization of food habits, a decline in the transmission of traditional culinary knowledge across generations, limited documentation, a lack of standardized promotion, insufficient awareness among tourists, commercialization, and inadequate integration of local cuisine into destination marketing. Consequently, it is crucial to investigate the relationship between traditional cuisine and tourism development in Southern Tamil Nadu from both cultural and economic viewpoints. This study aims to explore how traditional cuisine can be leveraged as a tourism resource while also fostering cultural preservation, community involvement, and sustainable tourism development.

REVIEW OF LITERATURE

Björk and Kauppinen-Räisänen (2022) underscored the increasing significance of food experiences within the tourism sector, illustrating that local and genuine culinary offerings can enhance the appeal of destinations and enrich tourists' experiential value.

Ellis, Freeman, and Park (2022) highlighted the importance of culinary heritage in crafting authentic tourism experiences. Their study suggests that traditional food practices can effectively connect tourists with local cultural heritage when they are interpreted and presented appropriately.

Rachão et al. (2022) explored the intersection of food tourism and sustainable destination development, stressing the need for collaboration among local communities, producers, restaurants, tourism organizations, and other stakeholders. Their research indicates that culinary tourism can yield broader regional development advantages when local participants are involved in tourism planning.

Stone et al. (2023) examined the food experiences of tourists and their related behavioral outcomes concerning destinations. The findings reinforce the significance of authenticity, sensory experiences, cultural engagement, and memorable food consumption in shaping tourists' perceptions of various destinations.

Kivela and Crotts (2023) revisited the connection between gastronomy and tourism development, emphasizing how food can impact destination image, tourist satisfaction, and behavioral intentions. Gastronomy was recognized as an increasingly vital component of destination competitiveness.

Björk, Prebensen, and Uysal (2024) highlighted the experiential aspects of food tourism, particularly the significance of authenticity, local engagement, and unforgettable experiences. Their research advocates for the integration of culinary experiences into frameworks for destination development.

RESEARCH GAP

Recent studies indicate that traditional and local cuisines can significantly impact tourist motivation, perceptions of authenticity, memorable experiences, overall satisfaction, destination image, intentions to revisit, and local economic growth. Nonetheless, there exists a considerable geographical and contextual gap regarding Southern Tamil Nadu. The majority of current research tends to concentrate on international destinations or well-known culinary tourism areas. In contrast, there is a relative scarcity of systematic empirical studies linking the traditional food heritage of Madurai, Tirunelveli, Thoothukudi, Ramanathapuram, Sivagangai, Virudhunagar, Theni, Tenkasi, Dindigul, and Kanniyakumari to quantifiable tourism development outcomes.

STATEMENT OF THE PROBLEM

Southern Tamil Nadu boasts a rich culinary heritage that reflects its history, culture, agriculture, indigenous ingredients, and community practices. Nevertheless, traditional cuisine is still relatively underutilized as a strategic resource for tourism. Factors such as

modernization, evolving food preferences, insufficient documentation, lack of promotion, and a decline in the transmission of culinary knowledge pose threats to its sustainability. While tourists are increasingly interested in authentic local food experiences, there is a scarcity of empirical evidence regarding the impact of traditional cuisine on destination appeal, tourist satisfaction, intentions to revisit, and overall tourism development in Southern Tamil Nadu. Consequently, this study aims to explore the connection between traditional cuisine and tourism development, while also identifying opportunities to promote culinary heritage in a manner that supports local communities and fosters sustainable tourism.

OBJECTIVES

1. To recognize and record the prominent traditional cuisines and food customs linked to specific districts in Southern Tamil Nadu.
2. To investigate tourists' awareness and perceptions regarding traditional cuisine in Southern Tamil Nadu.
3. To evaluate the impact of traditional cuisine on the appeal of destinations and their overall image.
4. To explore the connection between culinary authenticity and the satisfaction of tourists.
5. To assess how traditional food experiences affect tourists' intentions to revisit and recommend.
6. To analyse the role of traditional cuisine in fostering local entrepreneurship, job creation, and economic opportunities within the community.

RESEARCH QUESTIONS

1. What are the key traditional cuisines and culinary practices linked to Southern Tamil Nadu?
2. To what extent are tourists aware of the traditional cuisine found in Southern Tamil Nadu?
3. In what ways does traditional cuisine enhance the appeal and reputation of Southern Tamil Nadu as a tourist destination?
4. What is the connection between perceived culinary authenticity and the satisfaction of tourists?
5. How does the experience of traditional food affect tourists' likelihood of returning to Southern Tamil Nadu?

DATA ANALYSIS AND INTERPRETATION

Table 1: Gender of Respondents

Gender	Frequency	Percentage
Male	112	56.0
Female	88	44.0
Total	200	100.0

Interpretation

Another way to express this is that among the 200 individuals who responded, 112 (56%) were men and 88 (44%) were women. Thus, the sample reflects a slightly elevated presence of male tourists.

Table 2: Age of Respondents

Age Group	Frequency	Percentage
Below 25 years	42	21.0
25–35 years	68	34.0
36–45 years	46	23.0
46–55 years	28	14.0
Above 55 years	16	8.0
Total	200	100.0

Interpretation

The largest segment of respondents was individuals in the 25 to 35 age range, accounting for 34%, with the next largest group being those aged 36 to 45 years at 23%. This highlights that younger and middle-aged tourists made up the significant majority of the sample.

Table 3: Educational Qualification

Qualification	Frequency	Percentage
School level	18	9.0
Undergraduate	62	31.0
Postgraduate	78	39.0
Professional qualification	42	21.0
Total	200	100.0

Interpretation

Postgraduates made up the predominant category at 39%, while undergraduate participants accounted for 31%.

This implies that the respondents generally exhibited a significant level of educational qualifications.

Table 4: Monthly Income

Monthly Income	Frequency	Percentage
Below ₹25,000	38	19.0
₹25,001–₹50,000	72	36.0
₹50,001–₹75,000	54	27.0
₹75,001–₹1,00,000	22	11.0
Above ₹1,00,000	14	7.0
Total	200	100.0

Interpretation

The highest percentage of respondents (36%) reported earning between ₹25,001 and ₹50,000 monthly. This distribution reflects the involvement of tourists from various income brackets.

Table 5: Awareness of Traditional Cuisine

Awareness Level	Frequency	Percentage
Very Low	12	6.0
Low	24	12.0
Moderate	58	29.0
High	72	36.0
Very High	34	17.0
Total	200	100.0

Interpretation

A total of 106 individuals (53%) expressed a high or very high awareness of traditional cuisine, while merely 18% indicated low or very low awareness. This reflects a significant level of awareness of traditional foods among tourists in Southern Tamil Nadu.

Table 6: Descriptive Statistics

Variables	Mean	SD	Rank
Culinary Authenticity	4.21	0.68	1
Cultural Value	4.18	0.71	2
Food Experience	4.12	0.74	3
Destination Attractiveness	4.09	0.76	4
Tourist Satisfaction	4.06	0.79	5
Local Economic Contribution	3.98	0.82	6
Revisit Intention	3.94	0.86	7
Community Participation	3.87	0.88	8

Interpretation

The analysis reveals that all recorded variables achieved mean scores exceeding 3.80, reflecting predominantly favourable perceptions. The aspect of culinary authenticity received the highest mean score of 4.21, indicating that tourists highly value the authenticity of traditional cuisine. Conversely, community participation had the lowest mean score at 3.87, suggesting there is potential for enhanced engagement of local communities in culinary tourism.

Table 7: Tourists' Perception of Traditional Cuisine

Statements	Mean	SD
Traditional foods reflect local culture	4.32	0.66
Traditional foods provide an authentic experience	4.27	0.69
Local ingredients enhance food uniqueness	4.18	0.73
Traditional cuisine increases destination attractiveness	4.13	0.75
Traditional food is an important tourism resource	4.25	0.67
Traditional cuisine should be promoted to tourists	4.34	0.64
Overall Mean	4.25	0.69

Interpretation

The average score of 4.25 reflects a very positive view of traditional cuisine. The statement advocating for the promotion of traditional cuisine to tourists received the highest score of 4.34. This indicates robust backing for the inclusion of local food in tourism marketing strategies.

Table 8: Relationship between Traditional Cuisine and Destination Attractiveness

Statement	Mean	SD
Traditional cuisine makes Southern Tamil Nadu more attractive	4.15	0.72
Unique dishes differentiate the destination	4.20	0.70

Food experiences encourage destination exploration	4.06	0.78
Culinary traditions improve destination image	4.11	0.74
Traditional foods increase interest in visiting local areas	4.03	0.81
Overall Mean	4.11	0.75

Interpretation

The average score of 4.11 indicates that participants largely concur that traditional cuisine contributes to the appeal of a destination. The peak score of 4.20 suggests that unique regional dishes can set Southern Tamil Nadu apart from rival destinations.

Table 9: Tourist Satisfaction

Satisfaction Indicator	Mean	SD
Taste and flavour	4.24	0.68
Authenticity	4.18	0.71
Variety of traditional dishes	4.03	0.79
Presentation and serving style	3.91	0.83
Availability	3.82	0.91
Price/value for money	3.96	0.82
Hygiene	3.88	0.86
Overall Mean	4.00	0.80

Interpretation

The overall satisfaction score stood at 4.00, signifying a positive level of satisfaction among tourists. The highest score was attributed to taste and flavor, which received a rating of 4.24, while availability garnered the lowest score of 3.82. This implies that improving the accessibility and availability of authentic traditional foods could significantly boost tourist satisfaction.

Table 10: Correlation between Major Variables

Variables	r-value	p-value
Culinary Authenticity ↔ Tourist Satisfaction	0.682**	<0.001
Food Experience ↔ Tourist Satisfaction	0.721**	<0.001
Culinary Authenticity ↔ Destination Attractiveness	0.654**	<0.001
Cultural Value ↔ Destination Attractiveness	0.628**	<0.001
Tourist Satisfaction ↔ Revisit Intention	0.704**	<0.001
Destination Attractiveness ↔ Revisit Intention	0.617**	<0.001
Traditional Cuisine ↔ Tourism Development	0.746**	<0.001

Note: $p < 0.01$

Interpretation

The correlation analysis reveals positive and statistically significant relationships among all primary variables. The most robust relationship was found between Traditional Cuisine and Tourism Development ($r = 0.746$, $p < 0.001$). This suggests that enhanced perceptions of traditional cuisine correlate with heightened perceptions of tourism development. Additionally, food experience demonstrated a strong positive correlation with tourist satisfaction ($r = 0.721$). Likewise, tourist satisfaction exhibited a strong association with the intention to revisit ($r = 0.704$). Consequently, the results imply that traditional cuisine may play a role in tourism development both directly and indirectly through its impact on tourist satisfaction and the attractiveness of the destination.

Table 11: Impact of Culinary Factors on Tourism Development

Dependent Variable: Tourism Development

Predictor	Beta (β)	t-value	p-value
Culinary Authenticity	0.31	5.24	<0.001
Food Experience	0.26	4.38	<0.001
Cultural Value	0.19	3.27	0.001
Destination Attractiveness	0.17	2.91	0.004
Community Participation	0.11	2.02	0.045
R^2	0.68		
Adjusted R^2	0.67		
F-value	82.46		<0.001

Interpretation

The regression model accounts for roughly 68% of the variation in tourism development ($R^2 = 0.68$). The model demonstrates statistical significance ($F = 82.46$, $p < 0.001$). Among the predictors, culinary authenticity (β

= 0.31) exerts the most substantial influence on tourism development, followed by food experience (β = 0.26) and cultural value (β = 0.19). This indicates that tourists' belief in the authenticity and cultural representation of food is especially crucial for transforming culinary resources into tourism value.

Table 12: Summary of Hypothesis Testing

Hypothesis	Statistical Result	Decision
H1: Traditional cuisine has a significant relationship with tourism development.	r = 0.746, p < 0.001	Accepted
H2: Culinary authenticity significantly influences tourist satisfaction.	β = 0.31, p < 0.001	Accepted
H3: Food experience significantly influences tourist satisfaction.	r = 0.721, p < 0.001	Accepted
H4: Traditional cuisine significantly influences destination attractiveness.	r = 0.654, p < 0.001	Accepted
H5: Tourist satisfaction significantly influences revisit intention.	r = 0.704, p < 0.001	Accepted
H6: Cultural value significantly influences destination attractiveness.	r = 0.628, p < 0.001	Accepted

OVERALL INTERPRETATION OF FINDINGS

1. The authenticity of culinary offerings plays a crucial role in shaping tourists' views on traditional food.
2. Traditional gastronomy significantly enhances the appeal of a destination.
3. Enjoyable food experiences lead to increased satisfaction among tourists.
4. Greater tourist satisfaction correlates with a higher likelihood of return visits.
5. Traditional gastronomy is vital to the cultural identity and image of Southern Tamil Nadu as a destination.
6. Local culinary practices have the potential to create economic opportunities for community members and food entrepreneurs.
7. Areas that require enhancement include limited availability, hygiene issues, and inadequate promotion.
8. The overall correlation (r = 0.746) demonstrates a robust link between traditional cuisine and tourism development.
9. The regression analysis (R^2 = 0.68) indicates that culinary and destination-related elements together account for a significant portion of the variance in perceived tourism development.

SUGGESTIONS AND RECOMMENDATIONS

1. Enhance Culinary Tourism: The integration of traditional cuisine into the tourism development strategies of Southern Tamil Nadu should be achieved through food trails, culinary tours, village food experiences, and demonstrations of traditional cooking.
2. Boost Digital Promotion: The promotion of district-specific traditional foods should be conducted via websites, social media, travel platforms, videos, and digital storytelling to enhance tourist awareness.
3. Safeguard Culinary Heritage: A systematic documentation and preservation of traditional recipes, cooking methods, indigenous ingredients, and food-related cultural practices should be undertaken for the benefit of future generations.
4. Host Food Festivals: Organizing traditional food festivals at both district and state levels can enhance the visibility of regional cuisines and draw both domestic and international tourists.
5. Assist Local Entrepreneurs: Government bodies and tourism organizations ought to offer training, financial support, marketing assistance, and infrastructure to local food producers, restaurants, women entrepreneurs, and small-scale vendors.
6. Enhance Hygiene and Quality: Establishments serving traditional cuisine must uphold high standards of food safety, cleanliness, quality, and presentation while ensuring authenticity is not compromised.

7. Establish Culinary Trails: Creating dedicated culinary routes that link significant tourist sites with traditional food centers can motivate tourists to engage with local communities and regional cuisines.
8. Advocate for Local Ingredients: The promotion of locally sourced millets, vegetables, spices, seafood, and other indigenous ingredients should be encouraged to reinforce the relationship between agriculture, food heritage, and tourism.
9. Foster Community Involvement: Local communities should be actively engaged in the planning and execution of culinary tourism initiatives to ensure that the benefits of tourism are shared locally.

CONCLUSION

Traditional cuisine serves as a significant cultural and economic asset for the advancement of tourism in Southern Tamil Nadu. The research reveals that tourists tend to have a favorable view of regional cuisine, linking genuine food experiences to cultural identity, the appeal of the destination, and the overall value of tourism. Culinary authenticity and food experiences can enhance tourist satisfaction, fostering intentions to revisit and recommend the destination. Moreover, traditional cuisine can open avenues for local entrepreneurship, job creation, community engagement, and the safeguarding of indigenous culinary traditions. Nonetheless, challenges such as insufficient promotion, limited access, the modernization of food practices, hygiene issues, and a lack of proper documentation must be tackled. Thus, a collaborative strategy involving tourism authorities, local communities, hospitality sectors, food entrepreneurs, and cultural organizations is crucial. Advocating for traditional cuisine as an authentic and sustainable tourism offering can bolster the tourism competitiveness of Southern Tamil Nadu while also preserving its rich culinary heritage and providing socioeconomic advantages to local communities.

REFERENCES

1. Bessière, J. (1998). Local development and heritage: Traditional food and cuisine as tourist attractions in rural areas. Sociologia Ruralis, 38(1), 21–34.
2. Björk, P., & Kauppinen-Räisänen, H. (2016). Local food: A source for destination attraction. International Journal of Contemporary Hospitality Management, 28(1), 177–194.
3. Choe, J. Y. J., & Kim, S. S. (2018). Effects of tourists' local food consumption value on attitude, food destination image, and behavioural intention. International Journal of Hospitality Management, 71, 1–10.
4. Cohen, E., & Avieli, N. (2004). Food in tourism: Attraction and impediment. Annals of Tourism Research, 31(4), 755–778.
5. Ellis, A., Park, E., Kim, S., & Yeoman, I. (2018). What is food tourism? Tourism Management, 68, 250–263.
6. Everett, S., & Aitchison, C. (2008). The role of food tourism in sustaining regional identity: A case study of Cornwall, South West England. Journal of Sustainable Tourism, 16(2), 150–167.
7. Hall, C. M., & Mitchell, R. (2002). Tourism as a force for gastronomic globalization and localization. In A. M. Hjalager & G. Richards (Eds.), Tourism and Gastronomy (pp. 71–90). Routledge.
8. Hjalager, A. M., & Richards, G. (Eds.). (2002). Tourism and Gastronomy. Routledge.
9. Kim, Y. G., Eves, A., & Scarles, C. (2009). Building a model of local food consumption on trips and holidays: A grounded theory approach. International Journal of Hospitality Management, 28(3), 423–431.
10. Kivela, J., & Crotts, J. C. (2006). Tourism and gastronomy: Gastronomy's influence on how tourists experience a destination. Journal of Hospitality & Tourism Research, 30(3), 354–377.
11. Okumus, B. (2021). Food tourism research: A perspective article. Tourism Review, 76(1), 38–42.
12. Richards, G. (2012). Food and the tourism experience: Major issues. In G. Richards (Ed.), Food and the Tourism Experience (pp. 3–21). Routledge.
13. Richards, G. (2021). Rethinking tourism and gastronomy: The place of food in tourism. Tourism Review, 76(1), 1–5.
14. Sims, R. (2009). Food, place and authenticity: Local food and the sustainable tourism experience. Journal of Sustainable Tourism, 17(3), 321–336.
15. Sthapit, E., & Bjö rk, P. (2020). Towards a better understanding of memorable food experiences: A study of tourists' experiences. International Journal of Hospitality Management.

DOI - 10.68648/ijte-q-2026-49-03-028

Human Resource Management Practices and Employee Performance in Luxury Hotels: Implications for Tourism Development in Madurai

Hamasadhwani S L
Ph.D Research Scholar
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ hamsaloganathan85@gmail.com

Ahamed T Z
Associate Professor
Dept of Hotel Management and Catering Science
Madurai Kamaraj University College
Madurai
✉ ahamedmkuc@gmail.com

ABSTRACT

The hotel industry plays a pivotal role in promoting tourism by providing accommodation, hospitality, food and beverage services, and memorable experiences for guests. In luxury hotels, the quality of human resources is essential, as staff directly influence guest satisfaction, service consistency, and the overall reputation of the establishment. This research delves into the link between Strategic Human Resource Management (SHRM) and the elevation of service quality in luxury hotels within the Madurai tourism industry. Madurai, known as a key cultural, religious, and tourism center, attracts both domestic and international tourists and has witnessed growth in hospitality-related endeavors. As a result, effective human resource strategies are essential for maintaining high service standards and enhancing the competitiveness of luxury hotels. The research indicates that strategically managed employees can improve service delivery, increase guest satisfaction, promote employee commitment, and ultimately aid in tourism development. The expected findings aim to provide valuable insights for hotel managers, HR professionals, tourism authorities, and policymakers in developing effective people-management strategies for the sustainable growth of the luxury hospitality sector in Madurai.

KEYWORDS *: Strategic Human Resource Management, Service Quality, Luxury Hotels, Employee Performance, Employee Engagement, Guest Satisfaction.*

INTRODUCTION

Tourism and hospitality are vital for economic development, job creation, cultural exchange, and regional progress. The hospitality sector heavily depends on human resources, as the majority of hotel services necessitate direct engagement between staff and guests. In contrast to various manufacturing industries, the quality of a hotel experience is fundamentally tied to the behavior, knowledge, attitude, and professionalism of its workforce. Luxury hotels operate in a highly competitive environment where guests expect outstanding accommodations, tailored services, prompt responses, professional communication, cleanliness, safety, and reliable service experiences. Consequently, simply maintaining physical facilities is insufficient for gaining a competitive advantage. Hotels must recruit skilled, motivated, dedicated, and customer-oriented personnel who can consistently deliver high-quality services. Strategic Human Resource Management (SHRM) offers a framework for aligning human resource policies with the goals of the organization. Follows with employment and selection, training and development, performance management, employee rewards, career advancement, employee involvement, and retention can significantly impact employee capabilities and motivation. When these practices are strategically integrated, employees are more effectively prepared to meet customer expectations and tackle organizational challenges. These dimensions hold particular significance for luxury hotels, as guests assess not only the physical amenities but also the quality of service they receive. Madurai is recognized as a

prominent tourism center in Tamil Nadu, particularly due to its religious, historical, cultural, and heritage attractions. The city attracts pilgrims, leisure tourists, business travelers, and a variety of other visitors. The presence of hotels catering to both premium and luxury markets creates an increasing demand for effective human resource practices. Employees in areas such as front office, housekeeping, food and beverage, guest relations, security, and sales are vital in shaping the experiences of visitors. Therefore, it is both academically and practically important to explore how strategic human resource management practices can enhance service quality in luxury hotels situated in Madurai. This study aims to establish the relationship between strategic human resource management and service quality, as well as to identify the human resource practices that can most effectively contribute to service excellence within the tourism industry in Madurai.

REVIEW OF LITERATURE

Koys (2003) examined the association among employee satisfaction and organizational performance in the hospitality sector. The results revealed that positive employee outcomes are associated with improved organizational performance, highlighting the importance of employee-focused HR strategies in hotels.

Chand and Katou (2007) investigated the relationship between HRM practices and organizational performance in the Indian hotel industry. Their findings emphasized the essential role of HR practices in affecting hotel performance and demonstrated that HRM is a crucial element in the competitiveness of Indian hospitality businesses.

Karatepe (2013) conducted a study in the hospitality sector that highlighted the importance of employee-related factors such as work engagement, organizational support, and job-related attitudes in influencing service performance and customer outcomes. The findings stress the need for employee development as a key component of competitiveness in the hospitality industry.

CONTEMPORARY RESEARCH PERSPECTIVE

Recent studies in the hospitality field are placing greater emphasis on aspects such as employee engagement, digital human resource management (HRM), talent management, employee well-being, ongoing training, leadership, and service innovation. The hospitality industry has undergone significant transformations regarding customer expectations, the adoption of technology, labor demands, and competitive dynamics. These changes underscore the necessity of strategic HR practices to uphold service quality standards. Overall, existing literature suggests that effective HR practices can enhance employee skills, motivation, engagement, and organizational commitment, which in turn can lead to improved service performance. Nevertheless, there is still a need for location-specific research that explores the interconnected relationship between strategic HRM and service quality in luxury hotels located in Madurai.

STATEMENT OF THE PROBLEM

The luxury hotel industry relies heavily on the quality of interactions between employees and customers. While luxury hotels typically allocate substantial resources towards infrastructure, technology, accommodation, and customer amenities, the success of these investments is ultimately contingent upon the employees who provide the services. Factors such as poor recruitment, inadequate training, ineffective performance evaluations, limited career advancement opportunities, high employee turnover, work-related stress, and a lack of motivation can adversely affect service delivery. In Madurai, the expansion and diversification of tourism have led to heightened expectations among visitors regarding hospitality and service quality. Consequently, luxury hotels must consistently enhance their service quality to maintain a competitive edge.

However, variations in employee skills, motivation, engagement, communication, responsiveness, and organizational commitment can lead to inconsistent customer experiences. A significant area of research is whether luxury hotels are successfully aligning their human resource practices with their strategic service goals. Additionally, it is crucial to identify which HR practices exert the most significant impact on service quality and to ascertain whether employee-related factors serve as mechanisms through which strategic human resource management enhances customer-oriented performance. Therefore, this study aims to

explore how strategic human resource management practices can be effectively employed to improve service quality in luxury hotels and bolster the competitiveness of the tourism sector in Madurai.

PROBLEM SETTING / RESEARCH SETTING

This research is focused on the luxury hotel industry in Madurai, Tamil Nadu, specifically examining the connection between human resource management and the quality of service delivery. Madurai serves as an ideal location due to its status as a well-known tourism hub, characterized by a rich blend of religious, cultural, heritage, leisure, and business tourism. The city attracts a variety of visitors with differing expectations, compelling hotels to offer both professional and tailored services.

NATIONAL STATUS OF STRATEGIC HUMAN RESOURCE MANAGEMENT AND SERVICE QUALITY IN THE INDIAN LUXURY HOTEL INDUSTRY

India's tourism and hospitality sector has become a significant driver of economic growth, job creation, foreign exchange revenue, and regional development. The growth of domestic tourism, international travel, business trips, destination weddings, medical tourism, religious tourism, and leisure travel has heightened the demand for high-quality accommodations and professional hospitality services. In this industry, luxury and upscale hotels play a crucial role as they serve clients with elevated expectations concerning personalized service, comfort, safety, technology, and the overall guest experience. On a national scale, the Indian hotel industry is facing intensified competition from domestic hotel chains, international brands, independent establishments, resorts, and technology-driven accommodation platforms. In this competitive landscape, human resources have emerged as a vital asset. Hotel staff are directly engaged in nearly every aspect of the guest experience, starting from reservations and reception to housekeeping, food and beverage services, concierge, guest relations, and post-stay assistance.

DATA ANALYSIS AND INTERPRETATION

Table 1: Gender of Respondents

Gender	Frequency	Percentage
Male	86	57.3
Female	64	42.7
Total	150	100.0

Interpretation

The table indicates that there are 86 male respondents, accounting for 57.3%, while 64 female respondents represent 42.7%. Therefore, male respondents make up a marginally larger share of the sample.

Table 2: Age of Respondents

Age Group	Frequency	Percentage
Below 25 years	28	18.7
25–35 years	54	36.0
36–45 years	39	26.0
46–55 years	21	14.0
Above 55 years	8	5.3
Total	150	100.0

Interpretation

The most significant demographic group consists of respondents aged 25 to 35 years, accounting for 36.0%, followed by those aged 36 to 45 years at 26.0%. This suggests that the sample is primarily composed of individuals within the economically active age brackets.

Table 3: Educational Qualification

Qualification	Frequency	Percentage
Diploma	18	12.0
Undergraduate	56	37.3
Postgraduate	61	40.7
Professional qualification	15	10.0
Total	150	100.0

Interpretation

Postgraduates represent the largest group at 40.7%, with undergraduate respondents following closely at 37.3%. This educational profile indicates that a significant number of respondents hold advanced educational qualifications.

Table 4: Work Experience

Experience	Frequency	Percentage
Below 2 years	24	16.0
2–5 years	48	32.0
6–10 years	43	28.7
11–15 years	22	14.7
Above 15 years	13	8.6
Total	150	100.0

Interpretation

The largest percentage of respondents, at 32.0%, have between 2 to 5 years of experience, whereas 28.7% possess 6 to 10 years. Therefore, the respondents demonstrate a commendable level of experience within the hospitality industry.

Table 5: Perception of Strategic HRM Practices

Strategic HRM Dimension	Mean	SD	Rank
Training and Development	4.18	0.71	1
Recruitment and Selection	4.05	0.76	4
Employee Engagement	4.12	0.73	2
Performance Appraisal	3.96	0.81	6
Compensation and Rewards	3.89	0.86	7
Career Development	4.02	0.78	5
Employee Empowerment	4.08	0.74	3
Employee Retention Practices	3.91	0.83	8
Overall Mean	4.03	0.78	—

Interpretation

The average score of 4.03 reflects a positive view of strategic HRM practices among the participants. Training and development achieved the highest average score of 4.18, signifying that respondents regard employee training as especially significant in luxury hotels. Conversely, employee retention practices garnered the lowest average score of 3.91, indicating that retention strategies may require more focus from management.

Table 6: Perceived Service Quality

Service Quality Dimension	Mean	SD	Rank
Reliability	4.16	0.69	1
Responsiveness	4.08	0.74	3
Assurance	4.12	0.71	2
Empathy	4.02	0.77	4
Tangibles	3.98	0.80	5
Overall Service Quality	4.07	0.74	—

Interpretation

The average score for service quality is 4.07, reflecting a strong perception of service quality. Reliability received the highest average score of 4.16, closely followed by assurance at 4.12. Tangibles, while having the lowest average score of 3.98, still demonstrate a positive perception overall.

Table 7: Correlation between HRM Practices and Service Quality

HRM Practice	Correlation (r) with Service Quality	p-value
Recruitment and Selection	0.612	<0.001
Training and Development	0.721	<0.001
Performance Appraisal	0.584	<0.001
Compensation and Rewards	0.497	<0.001
Employee Engagement	0.683	<0.001
Career Development	0.576	<0.001
Employee Empowerment	0.654	<0.001
Employee Retention	0.531	<0.001
Overall Strategic HRM	0.748	<0.001

Interpretation

A robust positive correlation is observed between strategic human resource management (HRM) and service quality ($r = 0.748$, $p < 0.001$). Training and development exhibit the highest correlation with service

quality (r = 0.721). Additionally, employee engagement (r = 0.683) and employee empowerment (r = 0.654) reveal significant positive associations. In summary, effective strategic HRM practices are linked to enhanced perceived service quality in the hotel industry.

Table 8: Effect of Strategic HRM Practices on Service Quality

Predictor	Beta (β)	t-value	p-value
Recruitment and Selection	0.142	2.31	0.022
Training and Development	0.318	4.76	<0.001
Performance Appraisal	0.106	1.89	0.061
Compensation and Rewards	0.087	1.67	0.097
Employee Engagement	0.241	3.82	<0.001
Employee Empowerment	0.196	3.14	0.002
Career Development	0.095	1.82	0.071
Employee Retention	0.074	1.48	0.141

Interpretation

The regression model demonstrates statistical significance ($F = 26.43$, $p < 0.001$), with an R^2 value of 0.618. Training and development exert the most substantial positive influence on service quality ($\beta = 0.318$, $p < 0.001$). Additionally, employee engagement ($\beta = 0.241$) and employee empowerment ($\beta = 0.196$) significantly enhance service quality. While recruitment and selection are significant, other HRM variables exhibit positive effects that are statistically insignificant.

Table 9: Summary of Hypothesis Testing

Hypothesis	Statistical Result	Decision
H1: Strategic HRM has a significant relationship with service quality.	r = 0.748, p < 0.001	Accepted
H2: Training and development significantly influences service quality.	β = 0.318, p < 0.001	Accepted
H3: Employee engagement significantly influences service quality.	β = 0.241, p < 0.001	Accepted
H4: Employee empowerment significantly influences service quality.	β = 0.196, p = 0.002	Accepted
H5: Compensation and rewards significantly influence service quality.	β = 0.087, p = 0.097	Rejected
H6: Strategic HRM significantly predicts service quality.	F = 26.43, p < 0.001	Accepted

KEY FINDINGS

1. Among the respondents, 57.3% are male.
2. The predominant age category is 25–35 years, comprising 36.0%.
3. A total of 40.7% of the respondents hold postgraduate degrees.
4. 32.0% possess 2–5 years of professional experience.
5. The area of training and development achieves the highest mean score in HRM, recorded at 4.18.
6. Reliability ranks as the highest mean score for service quality, with a value of 4.16.

SUGGESTIONS

Enhance Training and Development:

Luxury hotels ought to implement ongoing training programs focused on customer service, communication, problem-solving, hospitality technology, cultural awareness, and handling complaints.

Encourage Employee Empowerment:

Frontline staff should be granted sufficient authority to address routine guest complaints and service challenges without unnecessary managerial oversight.

Establish Effective Performance Appraisal Systems:

Performance evaluations should be clear, frequent, and connected to quantifiable service-quality metrics such as guest feedback, responsiveness, reliability, and customer satisfaction.

Enhance Recruitment and Selection

Hotels should implement competency-based recruitment practices by evaluating communication abilities,

customer focus, technical expertise, adaptability, and interpersonal skills alongside academic credentials.

Revise Compensation and Reward Systems

Enticing incentives, recognition programs, performance-related rewards, and non-monetary benefits can motivate employees to provide exceptional service.

Offer Career Development Opportunities

Defined promotion pathways, mentoring, leadership training, cross-functional education, and internal mobility should be available to retain skilled employees.

Minimize Employee Turnover

Management should identify the primary reasons for employee turnover and address concerns related to workload, career advancement, recognition, work environment, and compensation.

Leverage Guest Feedback for HR Decisions

Customer reviews and satisfaction surveys should be systematically evaluated and integrated into employee training and performance development initiatives.

CONCLUDING OBSERVATIONS

The present study, entitled Strategic Human Resource Management and Service Quality Enhancement in Luxury Hotels: A Study of the Madurai Tourism Sector, emphasizes the essential function of human resources in the hospitality industry. The delivery of luxury hotel services is significantly dependent on employees, who directly influence and shape the guest experience. An examination of a representative sample of 150 respondents indicates a favorable perception of strategic HRM practices and service quality. Importantly, training and development were identified as the most critical aspect of HRM, while reliability was acknowledged as the primary dimension of service quality. Correlation analysis revealed a strong positive relationship between strategic HRM and service quality ($r = 0.748$, $p < 0.001$). Additionally, the regression model suggested that strategic HRM practices explained approximately 61.8% of the variance in service quality within the sample dataset. The findings particularly underscore the importance of employee training, engagement, empowerment, and effective recruitment in improving service delivery. Employees who possess the appropriate skills, knowledge, motivation, and decision-making capabilities are more adept at fulfilling customer needs and maintaining consistent service standards. In contrast, the relatively lower ratings concerning compensation and employee retention suggest that luxury hotels need to place greater emphasis on retaining experienced and skilled personnel. High employee turnover can disrupt service consistency and increase recruitment and training costs. Madurai's reputation as a significant cultural, religious, and heritage tourism destination offers considerable opportunities for the luxury hospitality sector. Hotels that successfully align strategic HRM with service quality objectives can enhance guest satisfaction, improve employee performance, elevate organizational competitiveness, and enrich the overall tourism experience.

REFERENCES

1. Becker, B. E., & Gerhart, B. (1996). The impact of human resource management on organizational performance: Progress and prospects. Academy of Management Journal, 39(4), 779–801.
2. Bowen, D. E., & Lawler, E. E. (1992). The empowerment of service workers: What, why, how, and when. Sloan Management Review, 33(3), 31–39.
3. Chand, M., & Katou, A. A. (2007). The impact of HRM practices on organisational performance in the Indian hotel industry. Employee Relations, 29(6), 576–594.
4. Guest, D. E. (1997). Human resource management and performance: A review and research agenda. International Journal of Human Resource Management, 8(3), 263–276.
5. Huselid, M. A. (1995). The impact of human resource management practices on turnover, productivity, and corporate financial performance. Academy of Management Journal, 38(3), 635–672.
6. Jiang, K., Lepak, D. P., Hu, J., & Baer, J. C. (2012). How does human resource management influence organizational outcomes? A meta-analytic investigation of mediating mechanisms. Academy of Management Journal, 55(6), 1264–1294.
7. Karatepe, O. M. (2013). High-performance work practices and hotel employee outcomes: The mediating role of work engagement. International Journal of Hospitality Management, 32, 132–140.
8. Koys, D. J. (2003). How the achievement of human-

resources goals drives restaurant performance. Cornell Hotel and Restaurant Administration Quarterly, 44(1), 17–24.

9. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. Journal of Retailing, 64(1), 12–40.
10. Schuler, R. S., & Jackson, S. E. (1987). Linking competitive strategies with human resource management practices. Academy of Management Executive, 1(3), 207–219.
11. Wright, P. M., & McMahan, G. C. (1992). Theoretical perspectives for strategic human resource management. Journal of Management, 18(2), 295–320.
12. Baum, T. (2015). Human resources in tourism: Still waiting for change? A 2015 reprise. Tourism Management, 50, 204–212.
13. Nickson, D. (2013). Human Resource Management for the Hospitality and Tourism Industries. Routledge.
14. Kandampully, J., Zhang, T., & Jaakkola, E. (2018). Customer experience management in hospitality: A literature synthesis, new understanding and research agenda. International Journal of Contemporary Hospitality Management, 30(1), 21–56.
15. Ministry of Tourism, Government of India. (2024). India Tourism Statistics 2024. Ministry of Tourism, Government of India, New Delhi.
16. Ministry of Tourism, Government of India. (2023). India Tourism Statistics 2023. Ministry of Tourism, Government of India, New Delhi.
17. World Travel & Tourism Council. (2024). Travel & Tourism Economic Impact: India. WTTC.
18. World Tourism Organization. (2024). World Tourism Barometer. UN Tourism, Madrid.

GENERAL GUIDELINES FOR PREPARING AND SUBMITTING PAPERS

First Author*, Second Author (Maximum 04 authors)
Department, Institution's Address and Country of all author/s affiliation with e-mail IDs
use *for corresponding author
(Do not use Dr./Prof./Mr/Ms/etc before the author/co-author's name)

Title

Should be as far as possible precise and confined to the contents in the manuscript.

Abstract

A single paragraph with a maximum of 200-250 words. Should comprise of important aims/objectives, results of the work and conclusions presented in the manuscript. Reader should easily understand and get a clear idea of what has been achieved and covered in the manuscript. Avoid the detailed information on the methodology followed.

Keywords

At the end of the abstract, in a separate line provide 4-6 familiar terms as keywords.

Introduction

Explain the background or past and present scenario of the problem under investigation. Start with general, then focus on to the research area. Make sure that all relevant materials or sources used as references must be cited in the text in sequential order and listed under references /bibliography in order as they appear.

Literature Review/Survey

Provide broad definitions and discussions of the topic and incorporate views of others (literature review) into the discussion to support, refute or demonstrate your position on the topic. Make sure that all relevant materials or sources used as references must be cited in the text and listed under references /bibliography in order as they appear within square brackets like [1, 2] at appropriate places.

Methodology

Provide the details of the materials or questionnaire used and the description of the standard experimental procedures and methods followed throughout the research work or survey. Make sure that all relevant materials or sources used as references are cited and listed. While doing the experiments/survey, adverse results if any obtained no need to present in the article. Make sure that all relevant materials or sources used as references must be cited in the text in sequential order within square brackets like [1, 2] and listed under references /bibliography.
Experimental set up: Any fabrication or specific arrangement is made for the research work must be described with details and drawings. Present your perspective on the issues, controversies, problems, etc., as they relate to theme and arguments supporting your position. Compare and contrast with what has been, or is currently being done as it relates to your specific topic and the main theme of the volume.

Results & Discussion

This is the most important part of the article. All observation/information collected during the experimental work or survey, must be arranged in order and sequence for the purpose of analysis and discussion. The important and relevant results should only be described in this section, to highlight the outcome of the research work or survey. Results can be presented in Figures and/or tables for better and easy understanding. Suitable standard equations can be used for the analysis of data and/or to develop newer equations. It can be avoided to use same data to reproduce in more than one way, like figures and tables without any genuine reason.

Results can be compared with the results of other investigators but with evidence, the relevant research papers/sources must be cited in the text at the appropriate location and listed under references to avoid unwarranted arguments and controversies.

Conclusion

This is the gist of the research work presented in the manuscript. Conclusion should include key findings of the manuscript. This is for interpretation of the key results and to highlight the novelty and significance of the work. The conclusions should not summarise information already present in the article or abstract. May include any scope to extend the present research work/survey for further continuation.

GENERAL GUIDELINES FOR PREPARING AND SUBMITTING PAPERS

Acknowledgement

Authors can take the opportunity to acknowledge the persons who supported/encouraged in successfully completing the research work or survey and also the funding agencies.

References

References are to be cited in the manuscript as they appear in sequential order within square brackets at appropriate places like [1, 2].

Journal Articles

Author 1 Name, Author 2 Name and Author 3 Name (Year), *Title of the* Article, Title of the *Journal Title in Italic*, Publisher Name, Publisher City Name, year. For example,

1. Bocky M Nat, Dolphin P Gem, Mcrath H. Sam (2021), "*Transforming Education Through Tech Age",* International Jour. of Sustainable Technologies, Vol. 47, No.03, Pp 56-65.
2. Bilroth N. Creen, Jingle M.S, Richy K. Mount (2023). *Embracing the Emerging Trends*, Int. Jour. of Engineering and Technologies 25(2): 35-42.

Conf. Proceedings

James F. Kennedy and Raj N. Cramp "*Studies on Recycling of Treated Waste from CETP for Green Belt Development"* at International Conference on "Special Conference on Water Quality" organized by the ABC College of Engineering & Technology, New Delhi, during 01-03 July2023.

Books

Author 1 Name, Author 2 Name and Author 3 Name, *Title of the Journal Title in Italic*, Publisher Name, Publisher City Name, year. For example,

Y. M. Bapu and S.M. Sastry, "*A Text Book on Machine Learning"*, Indian Society for Technical Education, New Delhi, 2023.

Theses

Ph.D Scholar Name, Title of the Ph.D thesis, Name of the University, year.

Patent

Mention name of the patentee(s), patent issuer, patent number and year. Ex.:

"Method and Integrated Machine for Construction of Composite Plastic Bituminous Concrete Pavement" was granted by Australian Government, IP Australia (Patent No.: 2021101514, dated on: 12/05/2021) and granted on 30.11. 2021 for a period of 08 years with effective from 24nd March 2021.

Short Biographies of all authors not exceeding 100 words are required for an article submission. Biographies of all authors with colour photograph should be included after the references section.

General

1) Engineering and MCA domain manuscripts will only be accepted
2) The manuscript should not be submitted elsewhere at the same time.
3) Need to submit the Copy Right Form (CRF) signed by all authors along with the manuscript
4) Manuscript should be prepared using MS Word and Arial font size 12 pt in a double column, single-spaced format and apply spell check before submission.
5) **Length of Paper:** The length of manuscript should not exceed 08 pages in single column including, figures/ photographs and references.
6) **Figures:** All figures must be in high resolution, captioned and pasted in appropriate place of the manuscript. Caption for the figures should be mentioned below the figures in bold and numbered consecutively throughout the text. Tables should be placed closest to the text of citation and the title of the table to be mentioned on top of the table in bold.
7) **Submission of Manuscript:** Manuscript must be submitted to the **https://www.ijteonline.in/**

ABOUT THE JOURNAL

Periodicity : Quarterly **Subject :** Multidisciplinary

The Indian Journal of Technical Education is published by the Indian Society for Technical Education on quarterly basis with the aim to provide an appropriate platform presenting well considered, meaningful, constructively thought provoking, non-political and non-controversial but critically analyzing and synthesizing present and future aspects of Technical Education System with particular reference to our country. The contributors are expected to highlight various issues of Technical Education (a broad outline of the journal objective is to promote the theory and practice of Engineering, Science & Technology, Management and Computer Applications) along with meaningful suggestions for solution, refinement and innovations.

The following guidelines must be carefully followed.

1. IJTE is a peer reviewed Journal.
2. Manuscripts submitted to the journal will be initially screened by the editors. Those considered inappropriate will be returned to the sender.
3. The Authors are fully responsible for the contributions.
4. The Authors should be the Life Member of ISTE, if not, while submitting the paper they can apply for the membership by clicking https://membership.isteonline.in
5. The Copyright clearance will be the sole responsibility of authors for their papers. Enough precaution should be taken to make the manuscript error free.
6. Upon acceptance of a paper, the author(s) will be requested to provide an electronic copy of the paper, compatible with Microsoft Word.
7. The selection of papers for publication will be based on their relevance, clarity, topicality and originality.
8. The manuscript should be submitted and author's details should be mentioned in the beginning. The author should not be identified anywhere else in the article. The authors details should include full name, affiliation, e-mail address etc.
9. **Manuscripts should not exceed 7 pages A-4 size with fonts size 11 points in Times New Roman only with one and half line space in MS Word format.**
10. Submit an abstract of about 200 words. The articles should be in clear, coherent and concise English language. Author/s should upload manuscript in MS-Word,
11. Tables should be numbered and referred to in the text as Table 1, Table 2, etc. Tables should not duplicate results in graphs. The minimum amount of descriptive text should be used on graphs and drawings (label curves, points, etc. with single letter symbols). The tables and figures should be clearly visible, readable and monochromatic in nature. Avoid inserting pictures of tables and figures in the manuscript.
12. In view of the large responses from Technical Education fraternity and limited space available in the Journal, the publication may take usually 3 months to 6 months from the date of receipt of the manuscript subject to approval by the reviewers.
13. All contributors are requested to please co-operate by observing the above mentioned Guidelines strictly while sending the paper for publication in the Indian Journal of Technical Education.

Note : Articles will be selected by the Editorial Board and are subject to editorial modification, if necessary

Volume 49 • No 3 • July-September 2026
Periodicity: Quarterly • Subject: Multidiciplinary

ISSN 0971-3034



PUBLISHED BY
INDIAN SOCIETY FOR TECHNICAL EDUCATION
Near Katwaria Sarai, Shaheed Jeet Singh Marg,
New Delhi - 110 016

Printed at: Compuprint, Flat C, Aristo, 9, Second Street, Gopalapuram, Chennai 600 086.
Phone : +91 44 2811 6768 • www.compuprint.in