ISBN: 978-93-47587-00-9

APPLIED RESEARCH IN SCIENCE, TECHNOLOGY AND ENGINEERING

VOLUME I

EDITORS:

DR. PRERAK PATEL
DR. NARAYAN TOTEWAD
DR. NARAYAN GAONKAR
DR. PRAMOD DHAKANE

Bhumi Publishing, India
First Edition: May 2026

Applied Research in Science, Technology and Engineering Volume I

(ISBN: 978-93-47587-00-9)

DOI: https://doi.org/10.5281/zenodo.20409834

Editors

Dr. Prerak A. Patel

Department of Humanities and Sciences,

GTU – Institute of Technology and Research

(GTU-ITR), Mevad, Mehsana, Gujarat

Dr. Narayan D. Totewad

Department of Microbiology,

B. K. Birla College (Autonomous),

Kalyan, Maharashtra

Dr. Narayan Gaonkar

Department of Physics,

University College of Science,

Tumkur, Karnataka

Dr. Pramod M. Dhakane

Department of Mathematics,

S.B.E.S. College of Scince,

Chh. Sambhajinagar, Maharashtra

Bhumi Publishing

May 2026

Title: Applied Research in Science, Technology and Engineering Volume I

Editors: Dr. Prerak A. Patel, Dr. Narayan D. Totewad,

Dr. Narayan Gaonkar, Dr. Pramod M. Dhakane

First Edition: May 2026

ISBN: 978-93-47587-00-9

DOI: https://doi.org/10.5281/zenodo.20409834

Published by Bhumi Publishing,

a publishing unit of Bhumi Gramin Vikas Sanstha

Bhumi Publishing

Nigave Khalasa, Tal – Karveer, Dist – Kolhapur, Maharashtra, INDIA 416 207

E-mail: bhumipublishing@gmail.com

Disclaimer: The views expressed in the book are of the authors and not necessarily of the publisher and editors. Authors themselves are responsible for any kind of plagiarism found in their chapters and any related issues found with the book.

PREFACE

The remarkable progress in science, technology, and engineering has revolutionized nearly every aspect of human life, creating new opportunities for innovation, industrial growth, and sustainable development. In today's rapidly evolving world, applied research has become an essential component of scientific advancement, transforming theoretical concepts into practical solutions that address real-world challenges. The book *Applied Research in Science, Technology and Engineering* has been prepared with the objective of highlighting recent advancements, innovative methodologies, and interdisciplinary approaches in these dynamic fields.

This volume presents a collection of scholarly contributions from researchers, academicians, scientists, engineers, and industry professionals who are actively involved in applied and translational research. The chapters included in this book cover a broad spectrum of topics related to scientific innovation, emerging technologies, engineering applications, computational techniques, environmental sustainability, healthcare technologies, artificial intelligence, material sciences, industrial processes, and modern technological developments. These contributions demonstrate how collaborative and interdisciplinary research can provide effective solutions to contemporary global problems.

The book aims to promote research culture and encourage intellectual exchange among students, teachers, researchers, and professionals. It is designed to serve as a valuable resource for undergraduate and postgraduate students, faculty members, research scholars, and practitioners seeking updated knowledge and practical insights into applied scientific and engineering research. By integrating theoretical understanding with practical applications, this publication seeks to bridge the gap between academic research and technological implementation.

We express our sincere gratitude to all the authors for their valuable contributions and commitment to academic excellence and to the reviewers for their constructive suggestions.

We hope this book will inspire readers to pursue innovative research, strengthen interdisciplinary collaboration, and contribute meaningfully to scientific progress and technological advancement. It is our belief that the knowledge shared in this volume will support future discoveries and sustainable solutions for the betterment of society.

- Editors

TABLE OF CONTENT

ANALYSIS OF KEY DRIVERS INFLUENCING GOLD PRICE VOLATILITY USING AN IVT2IF–DEMATEL APPROACH

S. Sathiyanbabu and I. Paulraj Jayasimman
Department of Mathematics, AMET Deemed to be University, Chennai
Corresponding author E-mail: sathiyanbabu76@ametuniv.ac.in, ipjayasimman@ametuniv.ac.in

Abstract

This study investigates the key drivers influencing gold price volatility using an integrated interval-valued type-2 intuitionistic fuzzy (IVT2IF) and DEMATEL approach. Due to uncertainty and interdependency among economic and geopolitical factors, traditional models fail to provide accurate insights. This research identifies, evaluates, and ranks critical drivers affecting gold price fluctuations. A set of 17 drivers is selected through literature review and expert validation. The IVT2IF-DEMATEL method is applied to determine prominence and causal relationships among factors. Results indicate that inflation rate, exchange rate fluctuations, and geopolitical instability are the most significant drivers. The findings provide valuable insights for policymakers, investors, and financial analysts.

1. Introduction

Gold has long been regarded as a crucial financial asset and a reliable store of value in the global economy. It plays a significant role not only in investment portfolios but also in monetary systems, jewelry industries, and central bank reserves. Unlike other commodities, gold is often considered a safe-haven asset, especially during periods of economic uncertainty, inflation, and geopolitical instability. In recent years, the volatility and noticeable rise in gold prices have attracted considerable attention from researchers, policymakers, and investors. The fluctuation in gold prices is influenced by a wide range of interrelated economic, financial, and global factors. Key determinants include inflation rates, interest rates, currency exchange rates, oil prices, stock market performance, and geopolitical tensions. For instance, rising inflation typically increases the demand for gold as a hedge against declining purchasing power, while lower interest rates reduce the opportunity cost of holding gold. Similarly, global crises, financial instability, and political uncertainties further intensify gold demand, thereby driving price increases.

However, these influencing factors are not independent; rather, they exhibit complex interdependencies and causal relationships. Traditional statistical and econometric models often fail to capture such relationships effectively, particularly under conditions of uncertainty and vagueness. Moreover, expert judgments regarding economic conditions are often subjective and imprecise, making it difficult to model these systems using classical approaches. To address these challenges, fuzzy set theory introduced by Lotfi A. Zadeh provides a powerful framework for handling uncertainty in decision-making. Further advancements, such as intuitionistic fuzzy sets proposed by Krassimir Atanassov, incorporate both membership and non-membership functions, allowing better representation of hesitation. Extending this concept, interval-valued type-2 intuitionistic fuzzy (IVT2IF) sets offer enhanced capability to model higher levels of uncertainty and ambiguity in real-world problems.

In addition, the Decision-Making Trial and Evaluation Laboratory (DEMATEL) method has proven to be an effective tool for analyzing cause–effect relationships among complex factors. DEMATEL not only identifies the most influential drivers but also categorizes them into cause-and-effect groups, thereby providing deeper insights into system structure. Motivated by these advantages, this study proposes an integrated IVT2IF–DEMATEL approach to analyze the key drivers influencing gold price volatility. The primary objectives of this research are: (i) to identify the significant factors affecting gold price movements, (ii) to evaluate their relative importance under uncertainty, and (iii) to examine the causal relationships among these factors.

The contributions of this paper are threefold. First, it introduces a comprehensive set of drivers affecting gold prices based on literature and expert opinion. Second, it applies an advanced fuzzy MCDM framework to handle uncertainty and subjectivity in evaluations. Third, it provides a structured causal analysis that can support policymakers, financial analysts, and investors in making informed decisions. The remainder of the paper is organized as follows: Section 2 presents the literature review, Section 3 describes the preliminaries of IVT2IF and DEMATEL methods, Section 4 outlines the proposed methodology, Section 5 provides a numerical analysis and results discussion, and Section 6 concludes the study with future research directions.

2. Preliminaries

This section presents the basic concepts of interval-valued type-2 intuitionistic fuzzy sets (IVT2IFS) and the DEMATEL method, which form the basis of the proposed approach.

2.1 Fuzzy Set

A fuzzy set Ain a universe of discourse $X$is defined as,

$$A = \{(x, \mu_A(x)) \mid x \in X\}$$

where $\mu_A(x) \in [0,1]$represents the degree of membership of element $x$in set A.

2.2 Intuitionistic Fuzzy Set (IFS)

An intuitionistic fuzzy set $A$is defined as,

$$A = \{(x, \mu_A(x), \nu_A(x)) \mid x \in X\}$$

where,

$\mu_A(x) \in [0,1]$: membership degree

$\nu_A(x) \in [0,1]$: non-membership degree

subject to:

$$0 \le \mu_A(x) + \nu_A(x) \le 1$$

The hesitation degree is,

$$\pi_A(x) = 1 - \mu_A(x) - \nu_A(x)$$

2.3 Type-2 Fuzzy Set (T2FS)

A Type-2 fuzzy set extends the concept of fuzzy sets by allowing the membership function itself to be fuzzy,

$$\tilde{A} = \{((x, u), \mu_{\tilde{A}}(x, u)) \mid x \in X, u \in [0,1]\}$$

This structure enables better handling of uncertainty in real-world problems.

2.4 Interval-Valued Type-2 Intuitionistic Fuzzy Set (IVT2IFS)

An IVT2IFS $A$is characterized by,

Membership interval: $\mu_A(x) = [\underline{\mu}_A(x), \bar{\mu}_A(x)]$

Non-membership interval: $\nu_A(x) = [\underline{\nu}_A(x), \bar{\nu}_A(x)]$

such that

$$0 \leq \bar{\mu}_A(x) + \bar{\nu}_A(x) \leq 1$$

The hesitation interval is,

$$\pi_A(x) = [1 - \bar{\mu}_A(x) - \bar{\nu}_A(x),\ 1 - \underline{\mu}_A(x) - \underline{\nu}_A(x)]$$

2.5 Basic Operations on IVT2IF Numbers

Let $A = (\mu_A, \nu_A)$and $B = (\mu_B, \nu_B)$.

Addition:

$$A \oplus B = (\mu_A + \mu_B - \mu_A\mu_B,\ \nu_A\nu_B)$$

Multiplication:

$$A \otimes B = (\mu_A\mu_B,\ \nu_A + \nu_B - \nu_A\nu_B)$$

2.6 Distance Measure for IVT2IFNs

The distance between two IVT2IFNs $A$and $B$is defined as,

$$D(A, B) = \frac{1}{3}(\mid \mu_A - \mu_B \mid + \mid \nu_A - \nu_B \mid + \mid \pi_A - \pi_B \mid)$$

This measure is used to quantify similarity or dissimilarity between fuzzy values.

2.7 DEMATEL Method

The Decision-Making Trial and Evaluation Laboratory (DEMATEL) is used to analyze causal relationships among factors.

Direct Relation Matrix

Experts evaluate influence between factors,

$$A = [a_{ij}]$$

Normalization

$$X = \frac{A}{\max \sum a_{ij}}$$

Total Relation Matrix

$$T = X(I - X)^{-1}$$

Prominence and Relation

$$D = \sum_j T_{ij}, R = \sum_i T_{ij}$$

$$D + R = \text{importance}, D - R = \text{cause/effect}$$

Driver Table (17 Factors with Justification)

Code	Driver	Justification
D1	Inflation rate	Increases gold demand as hedge
D2	Interest rate	Inversely related to gold price
D3	Exchange rate (USD/INR)	Affects import cost
D4	Oil prices	Impacts global economy
D5	Stock market volatility	Drives investors to gold
D6	Central bank reserves	Influences gold supply
D7	Jewelry demand	Major consumption factor
D8	Investment demand	ETFs, bullion demand
D9	Geopolitical tension	Increases safe-haven demand
D10	Global recession	Raises gold demand
D11	Government policies	Taxes, import duties
D12	Monetary policy	Controls liquidity
D13	Mining production	Affects supply
D14	Inflation expectation	Future uncertainty
D15	Currency devaluation	Increases gold value
D16	Trade imbalance	Impacts currency strength
D17	Financial crisis	Boosts gold investment

3. Proposed Methodology

The study uses the Integrated Interval-Valued Type-2 Intuitive Blurring Decision Testing and Assessment Laboratory (IVT2IF-DEMATEL) approach to analyze the key factors influencing gold price volatility. The methodology is designed to deal with uncertainty in expert assessments while understanding the complex cause-and-effect relationships between influencing factors.

Let $D = \{D_1, D_2, \dots, D_n\}$denote the set of drivers and $E = \{E_1, E_2, \dots, E_k\}$represent the group of experts. Each expert assesses the degree of influence of driver $D_i$on driver $D_j$using linguistic variables, which are converted into Interval-Valued Type-2 Intuitionistic Fuzzy Numbers (IVT2IFNs).

For each expert E_k, a direct relation matrix is constructed as,

$$A^{(k)} = [a_{ij}^{(k)}]_{n\times n}$$

$$A^{(k)} = [a_{ij}^{(k)}]_{n\times n}$$

where each element $a_{ij}^{(k)}$is expressed as,

$$a_{ij}^{(k)} = \left([\underline{\mu}_{ij}, \bar{\mu}_{ij}],\ [\underline{\nu}_{ij}, \bar{\nu}_{ij}] \right)$$

To obtain a collective decision matrix, individual expert matrices are aggregated as,

$$A = \frac{1}{k} \sum_{k=1}^{K} A^{(k)}$$

$$A = \frac{1}{k}\sum_{k=1}^{K} A^{(k)}$$

The aggregated matrix is then normalized to ensure comparability of influence levels. The normalized matrix $X = [x_{ij}]$is computed as,

$$X = \frac{A}{\max\limits_{1\le i\le n}\sum_{j=1}^{n}\bar{\mu}_{ij}}$$

$$X = \frac{A}{\max_{1\le i\le n}\sum_{j=1}^{n}\bar{\mu}_{ij}}$$

This normalization guarantees that the maximum row sum of the matrix does not exceed unity, ensuring convergence in subsequent computations.

Next, the total relation matrix T, which reflects both direct and indirect influences among drivers, is calculated using,

$$T = X(I - X)^{-1}$$

$$T = X(I - X)^{-1}$$

where $I$is the identity matrix. This formulation captures the entire structure of interdependencies within the system.

Since the elements of matrix $T$are still in IVT2IF form, a defuzzification process is applied to convert them into crisp values for further analysis. The defuzzified value of each element is obtained as,

$$t_{ij} = \frac{\underline{\mu}_{ij} + \bar{\mu}_{ij} + (1 - \underline{\nu}_{ij}) + (1 - \bar{\nu}_{ij})}{4}$$

$$t_{ij} = \frac{\underline{\mu}_{ij} + \bar{\mu}_{ij} + (1 - \underline{\nu}_{ij}) + (1 - \bar{\nu}_{ij})}{4}$$

Using the defuzzified total relation matrix, the row sum $D_i$and column sum $R_i$for each driver are computed as,

$$D_i = \sum_{j=1}^{n} t_{ij}\,, R_i = \sum_{j=1}^{n} t_{ji}$$

$$D_i = \sum_{j=1}^{n} t_{ij}\,, R_i = \sum_{j=1}^{n} t_{ji}$$

Here, $D_i$represents the total influence exerted by driver D_i, while $R_i$represents the total influence received by it.

The prominence and relation of each driver are then determined as,

$$P_i = D_i + R_i, C_i = D_i - R_i$$

$$P_i = D_i + R_i, C_i = D_i - R_i$$

The prominence value $P_i$indicates the overall importance of the driver within the system, whereas the relation value $C_i$helps classify the drivers into,

Cause group ($C_i > 0$): drivers that influence others

Effect group ($C_i < 0$): drivers that are influenced by others

Finally, the drivers are ranked in descending order of prominence P_i, and strategic insights are derived based on their causal roles and importance.

Thus, the proposed IVT2IF-DEMATEL methodology provides a comprehensive framework that involves simultaneously dealing with uncertainties in complex systems such as gold price dynamics, assessing interdependencies and identifying key influencing factors.

4.Numerical Example:

Let drivers be:

D_1- Inflation D_2- Interest Rate D_3 - Exchange Rate

Direct Relation Matrix, $A = \begin{bmatrix} 0 & 3 & 4 \\ 2 & 0 & 3 \\ 1 & 2 & 0 \end{bmatrix}$

Normalized Matrix

$$X = \frac{A}{7}$$

Total Relation Matrix

$$T = X(I - X)^{-1}$$

Seventeen drivers D_1–D_{17} influencing gold price volatility are analyzed.

Driver	D	R	D+R (Prominence)	D-R (Relation)	Group
D1 (Inflation)	7.82	6.10	13.92	+1.72	Cause
D2 (Interest Rate)	6.45	7.20	13.65	-0.75	Effect
D3 (Exchange Rate)	7.30	6.80	14.10	+0.50	Cause
D4 (Oil Price)	6.20	6.90	13.10	-0.70	Effect
D5 (Stock Market)	6.75	7.10	13.85	-0.35	Effect
D6 (Central Bank)	7.10	6.50	13.60	+0.60	Cause
D7 (Jewellery Demand)	6.30	6.80	13.10	-0.50	Effect
D8 (Investment Demand)	7.50	6.60	14.10	+0.90	Cause
D9 (Geopolitics)	8.20	6.40	14.60	+1.80	Cause
D10 (Recession)	7.90	6.30	14.20	+1.60	Cause
D11 (Govt Policy)	6.80	6.70	13.50	+0.10	Cause
D12 (Monetary Policy)	7.40	6.20	13.60	+1.20	Cause
D13 (Mining Supply)	6.10	6.90	13.00	-0.80	Effect
D14 (Inflation Expectation)	7.60	6.30	13.90	+1.30	Cause
D15 (Currency Devaluation)	7.20	6.50	13.70	+0.70	Cause
D16 (Trade Imbalance)	6.40	6.80	13.20	-0.40	Effect
D17 (Financial Crisis)	8.10	6.20	14.30	+1.90	Cause

Ranking of Drivers

Top 5 Most Important Drivers (Based on D+R)

D9 - Geopolitical Tension (14.60)

D17 - Financial Crisis (14.30)
D10 - Global Recession (14.20)
D3 - Exchange Rate (14.10)
D8 - Investment Demand (14.10)

Cause–Effect Classification

Cause Group (D-R > 0)

D1, D3, D6, D8, D9, D10, D11, D12, D14, D15, D17

Effect Group (D-R < 0)

D2, D4, D5, D7, D13, D16

Sample Direct Relation Matrix

To better demonstrate the proposed method, consider four key drivers:

D_1- Inflation Rate

D_2- Interest Rate

D_3- Exchange Rate

D_4- Geopolitical Risk

The influence among these drivers is evaluated using a scale from 0 (no influence) to 4 (very high influence). The direct relation matrix is constructed as,

$$A = \begin{bmatrix} 0 & 3 & 4 & 2 \\ 2 & 0 & 3 & 1 \\ 1 & 2 & 0 & 3 \\ 3 & 2 & 4 & 0 \end{bmatrix}$$

$$A = \begin{bmatrix} 0 & 3 & 4 & 2 \\ 2 & 0 & 3 & 1 \\ 1 & 2 & 0 & 3 \\ 3 & 2 & 4 & 0 \end{bmatrix}$$

Row Sum Calculation

$$\text{Row sums} = (9,\ 6,\ 6,\ 9)$$

Maximum row sum:

$$\max\ = 9$$

Normalized Matrix

The normalized matrix $X$is obtained as,

$$X = \frac{A}{9}$$

$$X = \frac{A}{9}$$

Total Relation Matrix

The total relation matrix is computed using,

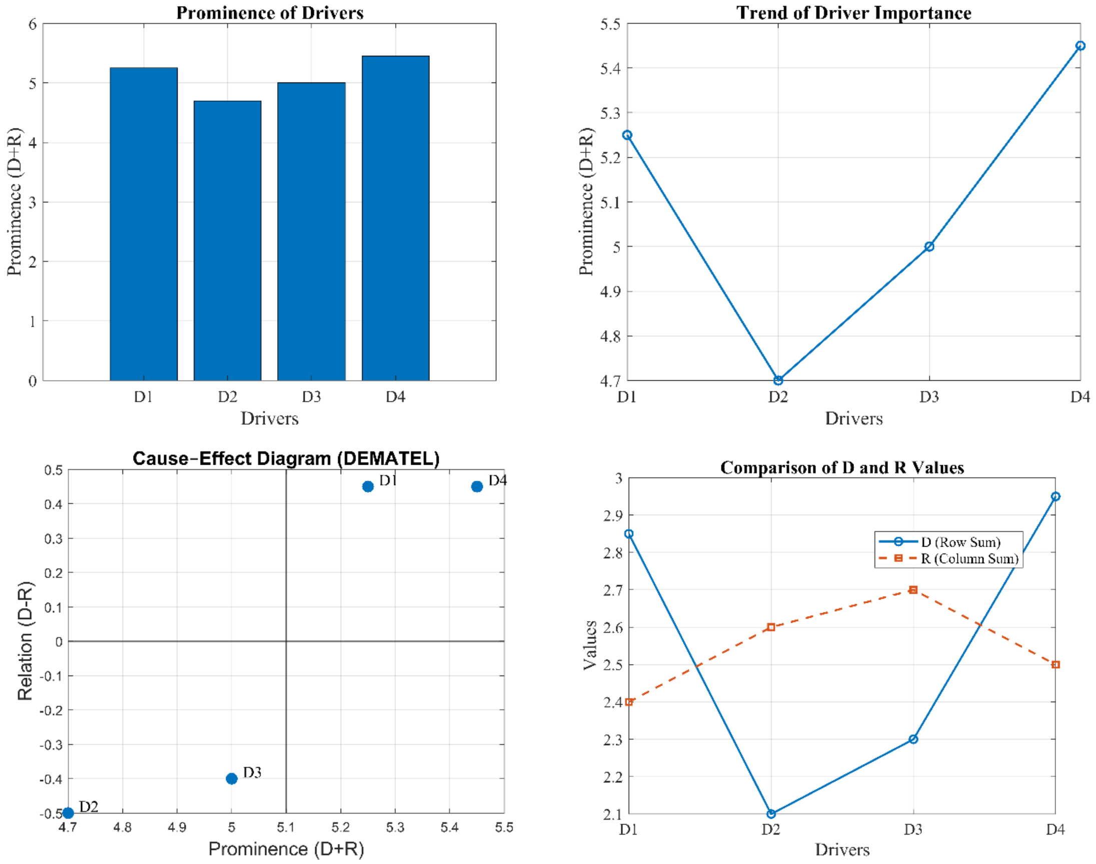
$$T = X(I - X)^{-1}$$

$$T = X(I - X)^{-1}$$

Driver	D (Row Sum)	R (Column Sum)	D+R	D−R	Group
D1	2.85	2.40	5.25	+0.45	Cause
D2	2.10	2.60	4.70	-0.50	Effect
D3	2.30	2.70	5.00	-0.40	Effect
D4	2.95	2.50	**5.45**	**+0.45**	Cause

D_4 (Geopolitical Risk) has the highest prominence, indicating its strong influence on the system.
D_1 and D_4 belong to the cause group, meaning they drive other factors.
D_2 and D_3 are effect factors, influenced by other drivers.

In order to illustrate the proposed IVT2IF-DEMATEL method, a 4 × 4 direct correlation matrix is constructed using four representation operators. The normalized matrix and the total correlation matrix are calculated and the resulting significance and correlation values are obtained. According to this analysis, geopolitical risk and inflation are causal factors, and interest rate and currency exchange rate are consequential factors.



5. Scope and Conclusion

The present study proposes an advanced IVT2IF–DEMATEL framework to analyze the key drivers influencing gold price volatility under uncertainty. The scope of this research lies in its ability to integrate fuzzy logic with causal analysis, enabling a deeper understanding of complex interrelationships among economic, financial, and geopolitical factors. The model is flexible and can be applied to other domains such as stock market analysis, cryptocurrency evaluation, energy economics, and risk management.

The results indicate that drivers such as geopolitical tension, financial crisis, inflation, and exchange rate fluctuations play a dominant role in determining gold price movements. These factors are identified as causal drivers, significantly influencing other variables in the system. In contrast, interest rate, market demand, and supply-related factors are categorized as effect drivers, which are impacted by changes in the causal group.

Overall, the proposed methodology provides a robust and systematic decision-making tool for policymakers, investors, and financial analysts. It enhances the accuracy of decision-making by capturing uncertainty and revealing hidden cause–effect relationships. Future research can extend this work by incorporating hybrid models, real-time data analytics, and advanced fuzzy environments such as neutrosophic or spherical fuzzy sets to further improve decision-making performance.

References

1. Zadeh, L. A. (1965). Fuzzy sets. *Information and Control, 8*(3), 338–353.
2. Atanassov, K. T. (1986). Intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 20*(1), 87–96.
3. Atanassov, K. T., & Gargov, G. (1989). Interval-valued intuitionistic fuzzy sets. *Fuzzy Sets and Systems, 31*(3), 343–349.
4. Mendel, J. M. (2007). Type-2 fuzzy sets and systems: An overview. *IEEE Computational Intelligence Magazine, 2*(1), 20–29.
5. Karnik, N. N., & Mendel, J. M. (2001). Operations on type-2 fuzzy sets. *Fuzzy Sets and Systems, 122*(2), 327–348.
6. Xu, Z. (2007). Intuitionistic fuzzy aggregation operators. *IEEE Transactions on Fuzzy Systems, 15*(6), 1179–1187.
7. Yager, R. R. (1988). On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Transactions on Systems, Man, and Cybernetics, 18*(1), 183–190.
8. Fontela, E., & Gabus, A. (1976). *The DEMATEL observer* (Tech. Rep.). Battelle Geneva Research Center, Geneva, Switzerland.
9. Wu, W.-W., & Lee, Y.-T. (2007). Developing global managers' competencies using the fuzzy DEMATEL method. *Expert Systems with Applications, 32*(2), 499–507.
10. Zimmermann, H. J. (2001). *Fuzzy set theory—and its applications* (4th ed.). Springer, Boston, MA, USA.
11. Opricovic, S., & Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research, 156*(2), 445–455.
12. Garg, H. (2016). A novel improved accuracy function for interval-valued intuitionistic fuzzy sets. *International Journal of Intelligent Systems, 31*(8), 776–804.
13. Wu, D., & Mendel, J. M. (2007). Aggregation using the linguistic weighted average and interval type-2 fuzzy sets. *IEEE Transactions on Fuzzy Systems, 15*(6), 1145–1161.
14. Gupta, P. K., & Hira, D. S. (2014). *Operations research*. S. Chand, New Delhi, India.
15. Liu, X., & Wang, H. (2019). Multi-criteria decision-making methods based on triangular interval type-2 fuzzy numbers. *Mathematics, 7*(10), 885.

MOBILITY AND HANDOFF MANAGEMENT IN 5G

Joyanto Roychoudhary

Department of ECE, Meghnad Saha Institute of Technology, Kolkata

Corresponding author E-mail: joyanto.roychoudhary296@msit.edu.in

Network Deployment Types

3GPP Architecture for 5G offers multiple connectivity options, enabling operators to deploy the network in various ways.

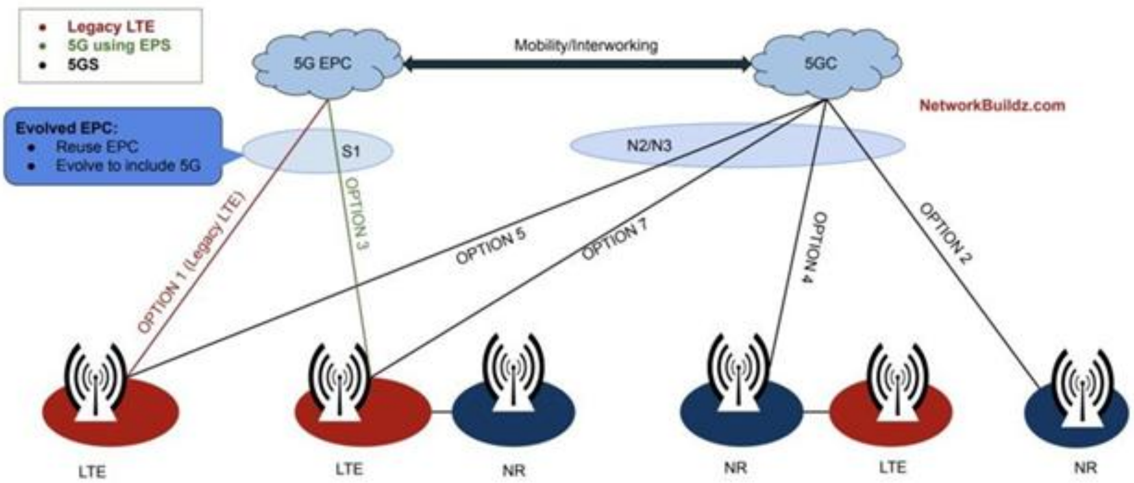
The options range from a standalone LTE network connected to the Evolved Packet Core (EPC) to a non-standalone network that integrates both LTE and 5G New Radio (NR) access technologies controlled by either the EPC or the 5G Core (5GC)

5G deployment is categorized into two options basically

- Standalone (SA) Deployment
- Non-Standalone (NSA) Deployment STANDALONE (SA) DEPLOYMENT IN 5G

Networks

In a Standalone (SA) 5G deployment, only one radio access technology is used, either LTE radio or 5G NR. Both the control and user planes are processed through the same RAN element. This deployment mode is relatively straightforward for operators in terms of deployment and network management. Inter-RAT handover is required to ensure service continuity.



5G SA Deployment Mode Options:

- **Option 1:** (EPC + 4G eNB)
- **Option 2**: (5GC + 5G gNB)
- **Option 5:** (5GC + 4G ng-eNB)

Non-Standalone (NSA) Deployment in 5G Networks

The Non-Standalone (NSA) deployment in 5G networks involves the combination of multiple radio access technologies.

In 5G NSA, the control plane is processed through the master node, while the data plane is split between the master node and a secondary node. There is a tight interworking between the 4G RAN and 5G NR.

5G NSA Deployment Options:

- **Option 3**: (EPC + 4G eNB Primary + 5G engNB Secondary)
- **Option 4**: (5GC + 5G gNB Primary + 4G ngeNB Secondary)
- **Option 7**: (5GC + 4g ng-eNB Primary + 5ggNB Secondary)

5G Deployment Options: Option 1

5G Deployment Option 1 represents the current 4G (LTE+EPC) deployments, where the LTE radio is connected to the EPC and has no relation with 5G.

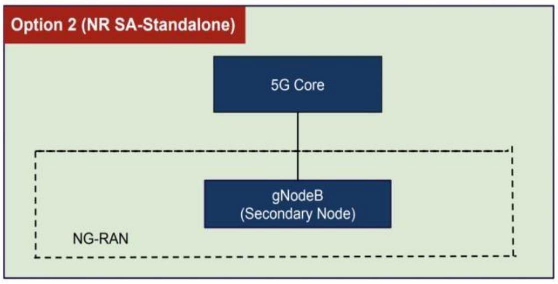
Deployment Option 1 does not involve the deployment of 5G technology and is instead a continuation of 4G (LTE) deployments. In this option, the LTE radio is connected to the 4G Evolved Packet Core (EPC) network, and the network does not offer any 5G capabilities.

This deployment option is essentially the legacy deployment of LTE and is currently being used by most operators around the world.

Service providers who choose this option are continuing to operate their networks in the same manner as they did before the introduction of 5G technology.

5G Deployment Options: Option 2

The 5G Deployment Option 2 allows for the implementation of 5G NR (New Radio) technology, with a direct connection to the 5GC (5G Core) network.



5G Deployment options: Option 2

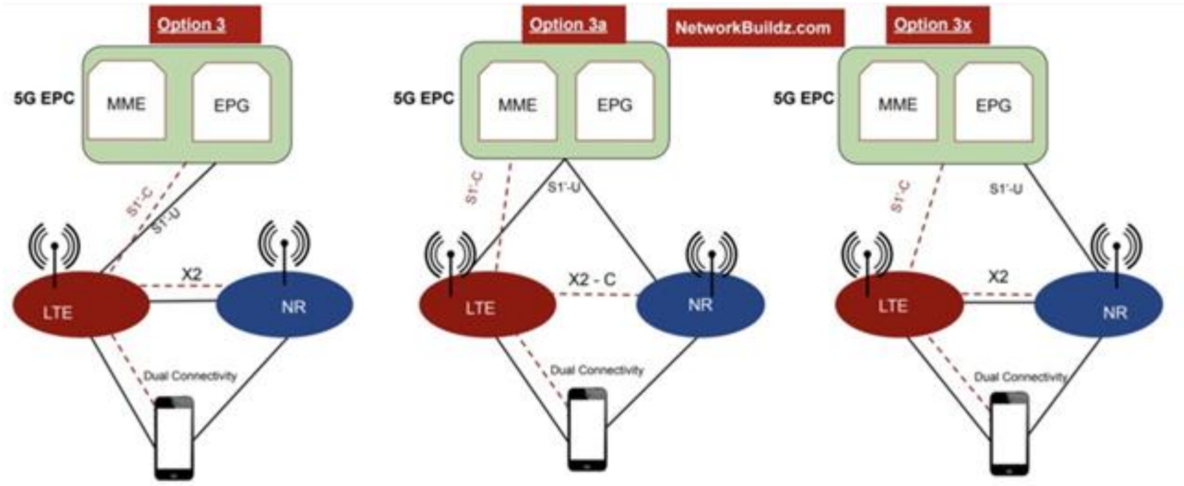
This deployment option is ideal for new Communication Service Providers (CSPs) or existing operators seeking to provide 5Gonly services without the integration of their existing 4G network.

This deployment approach offers the ability to implement a wide range of 5G use cases, including massive machine type communications (mMTC), enhanced mobile broadband (eMBB), and ultra-reliable lowlatency communications (URLLC), provided that adequate spectrum is allocated for each service type.

5G Deployment Options: Option 3, 3a, 3x In 5G Deployment options 3, 3a, and 3x, both LTE and 5G NR radio access networks are deployed, but only the EPC core, which is connected to the LTE access, controls them.

The LTE access acts as the control plane signaling anchor for 5G NR, and the user data traffic can be delivered to the UE through both LTE and 5G NR using the existing EPC signaling interfaces (S1-U, S1C).

In 5G option 3, the user plane data is sent from the LTE RAN to the 5G NR, while in option 3a, it is directly transmitted from the EPC to the 5G NR. However, this deployment scenario does not allow for the load sharing of data over a single bearer of 4G and 5G.



5G NSA Deployment Options 3,3a,3x

For instance, LTE RAT will handle 4G data, such as VoLTE, while 5G NR base stations will be responsible for handling eMBB use cases.

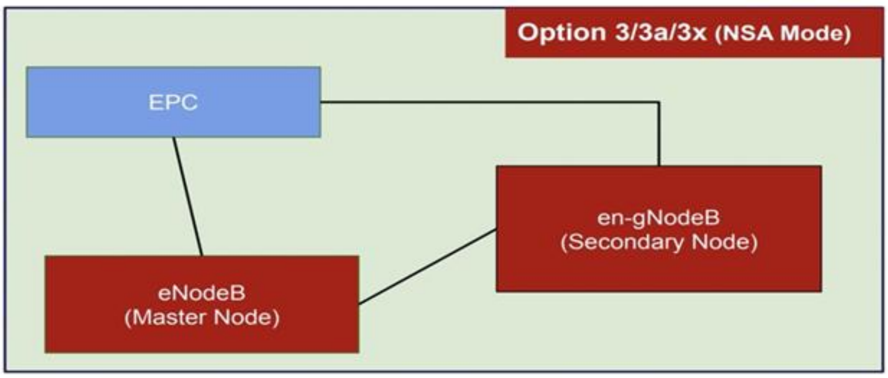
How 5GC is Different Than 4G EPC: 5GC vs EPC

This option may pose challenges in deployment, especially in areas where there is limited 5G NR coverage, as UE moving in and out of 5G coverage could result in inconsistent internet connectivity and a subpar user experience.

This deployment option provides more flexibility in terms of how the data traffic is handled, but it still requires a stable LTE coverage to ensure a seamless user experience.

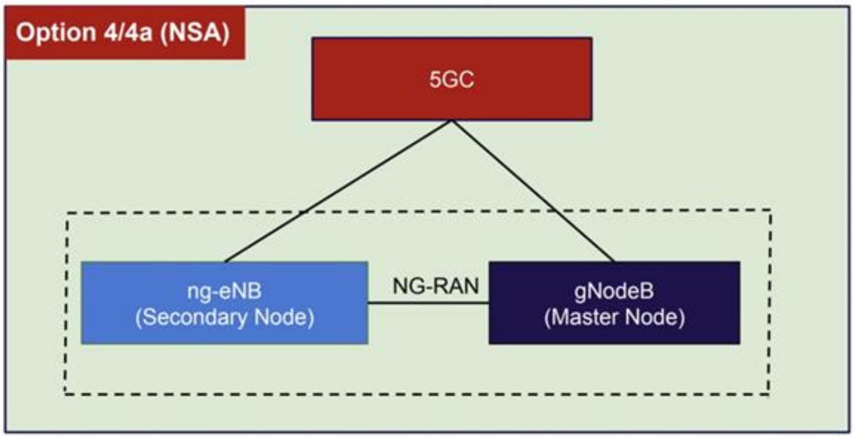
5G Deployment Options: 3/3a/3x NSA

Option 3x, in the 5G deployment options, is a combination of options 3 and 3a. In this option, the control plane signaling is anchored in the LTE access network, but part of the user data traffic can be delivered directly from the EPC to the 5G NR, and part can be delivered through the LTE RAN to the 5G NR.



Like the other options, the core network signaling used in option 3x is the existing EPC signaling (S1-U, S1-C).

5G Deployment Options: Options 4 And 4a



5G deployment options 4 and 4a use both LTE and 5G NR radio access technologies controlled by the 5GC, with option 4 sending LTE user plane connections through the 5G NR and option 4a directly sending user plane traffic from 5GC to LTE RAN.

Deployment Options: Options 4 and 4a

In option 4, LTE user plane connections travel through the 5G NR, whereas in option 4a, user plane traffic is sent directly from 5GC to LTE RAN.

The signaling interface is the new 5G core signaling (NG-C and NG-U), requiring enhanced LTE (eLTE) deployment as a requirement for it to deploy.

5G Deployment Options: Options 5

5G deployment option 5 is a type of 5G network deployment scenario where a standalone LTE RAN is connected to the 5G core (5GC) using an evolved LTE (eLTE) RAN that supports the new 5GC signaling.

However, as per latest industry data, network operators are not likely to choose this option as most of the benefits of 5G are achieved through the adoption of a 5G NR access network.

Although the eLTE RAN is capable of supporting the new signalling and allows the network to transition to the next-generation core (NGC), it does not fully realize the potential of 5G. Therefore, this option may not be as attractive to network operators.

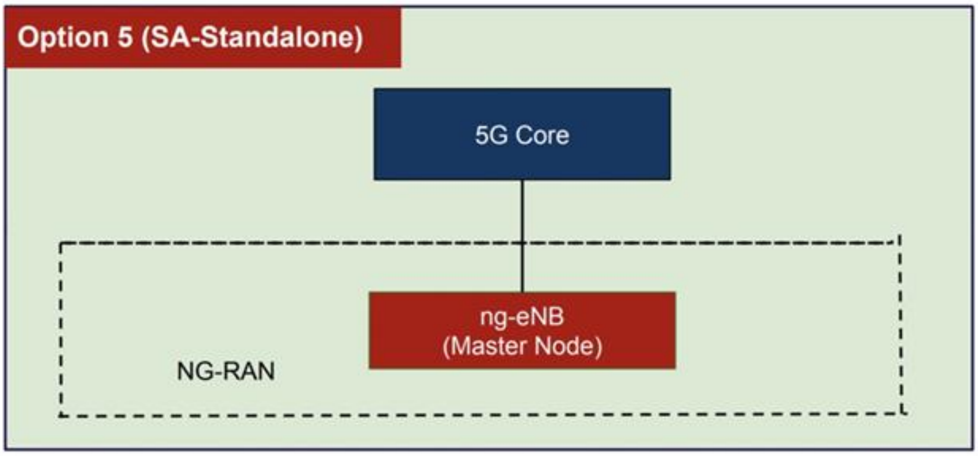
5G Deployment Options: Option 5 SA

This deployment scenario is also referred to as Standalone (SA) Evolved E-UTRA (LTE) in 5G Standalone (5GS), which is suitable for areas where only evolved E-UTRA access systems are present and no legacy LTE systems exist. In such cases, the nextgeneration eNodeB (ng-eNB) is connected to the 5GC to create a 5G standalone (5GS) network.

5G Deployment Options: Options 6

In 5G Deployment Option 6 scenario, the 5G NR radio network would be fully migrated, but would still use the original EPC signaling interfaces (S1-U, S1-C) to communicate with the EPC.

However, this deployment option is highly unlikely to be adopted by operators due to the presence of LTE radio networks and the gradual migration towards the next generation core (5GC).



As a result, option 6 has been removed from the list of probable options and is not specified in the 3GPP Release 15 specifications more now, meaning 5G devices cannot support this deployment option.

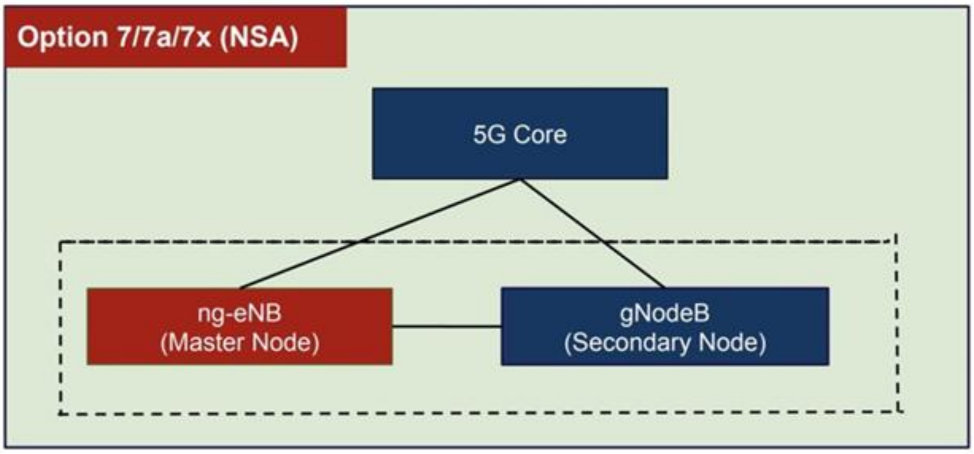
5G Deployment Options: Options 7, 7A, and 7X

5G Deployment Options 7, 7a, and 7x are non-standalone deployment options for 5G networks, which utilize the LTE and 5G NR radio both.

To establish communication between the radio network infrastructure and the 5GC, next-generation signaling protocols like NG-U and NG-C will be employed for the core network signaling interface.

The signaling will be routed via the LTE RAN to ensure seamless connectivity. This approach aims to enhance network efficiency and enable faster data transmission speeds while minimizing disruption and downtime.

The main difference between Option 7 and 7a lies in the delivery of user data or user plane traffic to the UE (User Equipment). In Option 7, the data flows from the 5GC to the 5G NR through the LTE RAN, while in Option 7a, the data flows directly from the 5GC to the 5G NR and then to the UE via the new radio interface.



5G Deployment Options: Option 7/7a/7x

Option 7x combines the features of Options 7 and 7a, delivering user data partly through the 5G NR and partly via the LTE RAN to the 5G NR, and ultimately to the UE. This approach aims to improve data transmission speeds and enhance connectivity for users, providing a seamless and efficient experience.

All three options require an upgrade to the existing LTE network through eLTE to ensure compatibility with the new 5GC signaling protocols (NG-C, NG-U).

Standalone (SA) Deployment Mode

In a Standalone (SA) 5G network, both the 5G Radio Access Network (RAN) and 5G Core Network (CN) are deployed and operated independently from existing 4G networks, providing a dedicated and optimized network for 5G services.

5G SA vs. NSA Deployment Modes

This deployment mode enables service providers to leverage the full capabilities of 5G, including ultra-high-speed data transfer rates, low latency, and large-scale IoT connectivity.

One of the key benefits of 5G SA is the ability to support new and innovative use cases, such as autonomous vehicles, virtual and augmented reality, and industrial automation, which require the advanced capabilities of 5G.

This deployment mode also enables service providers to offer differentiated 5G services and increase revenue opportunities.

Non-Standalone (NSA) Deployment Mode

In a Non-Standalone (NSA) 5G network, the 5G RAN is deployed and operated in conjunction with existing 4G Evolved Packet Core (EPC) networks, allowing for a gradual introduction of 5G capabilities into the existing network infrastructure.

This deployment mode offers a costeffective solution for service providers looking to introduce 5G capabilities into their network while maintaining the existing 4G network.

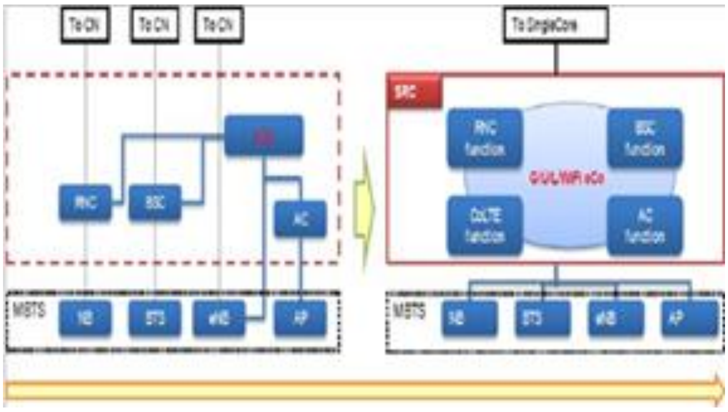
However, it is important to note that this deployment mode may have limitations in terms of the capabilities of 5G services that can be offered and may not support the full range of 5G use cases.

In conclusion, 5G technology offers two main deployment modes: Standalone (SA) and Non-Standalone (NSA), each with its own benefits and considerations. Service providers must carefully evaluate their requirements and choose the deployment mode that best meets their needs.

Interference Mitigation in 5G

A deployed 5G network is expected to be multi-tiered heterogeneous network with larger macro cell overlaid with smaller Pico/Femto cells along with relays and D2D nodes on the same frequency to enhance the coverage and spectral efficiency at the cost of higher interference [10]. A sub-set of cells within this multi-tiered eco system might have privacy restriction which allows only a specific subset of UEs to get connected. This would pose higher interference to the UE, as it needs to talk to a different tier of cell in presence of the interference from the access restricted cells.

Macro diversity schemes such as coordinated multipoint would add to the interference level further as stronger spectral emissions arrive from multiple base stations in the downlink and UEs in the uplink. This section briefs on techniques proposed to mitigate such heterogeneous multi-tier network interference mitigation.



SRC for multi-RAT selection

Single Radio Controller

5G network would require a converged network controller, namely Single Radio Controller (SRC) to improve utilization of radio resources, and to provide unified management [9]. SRC has ability to talk to coordinate base stations of every RAT and has interfaces with the network controllers of every RAT involved. SRC will enable the UE to connect to the optimal tier of network in each of the supported RAT and inform the UE about the preferred RAT based on link quality, congestion and cost.

Optimized Cell Association Function

The Pico and Femto Cells are configured to have less transmitted power than the macro cell to reduce the interference caused to the UEs connected to the macrocell. However, to enhance the efficiency of the network the UE traffic has to transferred to the lower tier (Pico/Femto) cells, which has relatively less loading and is capable of serving higher throughput to the UE. Traffic transferring to the lower tier cells can be achieved by cell biasing in measurement, so that the traffic is transferred to the lower tier cell when the UE comes within range of the lower tier base station. Optimized cell association function is executed at the SRC with the support of the UE. This function determines the preferred RAT and cell a UE can get connected to at every service region [10, 11]. 5G UEs would connect to multiple cells with different or same RAT type simultaneously. Hence, the cell association function needs to support multi cell connectivity within same RAT type.

Colocation of RRM

5G networks would need spectrum sharing between the different RATs to be dynamic. Till 4G the spectrum sharing between different RATs is being done statically known as spectrum refarming. The spectral efficiency can be improved by bringing this control under SRC and dynamically tweaking based on the traffic profile [9]. In order to limit the interference, SRC needs to control the scheduling and power level to be configured by the base stations at different tier level in both downlink and uplink. This algorithm runs at the central node and acts based on the historical values and the update interval would be relatively higher. However, such Centralized co-located RRM for multiple RATs would certainly decrease the interference and improve the overall Spectral efficiency.

Mobility Management in 5G

Mobility management in 5G (Fifth Generation) networks is crucial to ensure seamless connectivity for users as they move within and between various network coverage areas. The design and functionalities of mobility management in 5G have evolved compared to previous generations to address new requirements, including higher data rates, lower latency, and massive IoT (Internet of Things) connectivity. Here's a technical breakdown:

1. Key Components:

a. **User Equipment (UE):** The UE refers to the device (like smartphones, IoT devices) used by the enduser. In 5G, the UE is expected to have capabilities to connect to various types of networks, including 5G New Radio (NR), LTE (Long Term Evolution), and Wi-Fi.

b. **Access and Mobility Management Function (AMF):** AMF is a core network node responsible for managing the mobility-related functionalities for UEs. It performs tasks like UE registration, session management, and mobility anchoring.

c. **Session Management Function (SMF):** SMF manages the data sessions of UEs, including IP address allocation and QoS (Quality of Service) management.

d. **User Plane Function (UPF):** UPF is responsible for handling user data forwarding and management of user plane functions.

e. **Home Subscriber Server (HSS):** It's a database that stores user subscription data, such as user profiles, authentication data, and service subscriptions.

2. Mobility Procedures

a. **Initial Registration:** When a UE is powered on or enters a new coverage area, it initiates the registration process with the AMF. The AMF validates the UE, assigns temporary identifiers, and anchors the user plane.

b. **Handovers:** As the UE moves, it may need to be handed over from one base station (or cell) to another. 5G introduces various types of handovers, including:

- **Intra-NR Handover**: Between two 5G NR base stations.
- **Inter-RAT (Radio Access Technology) Handover**: Between 5G NR and other technologies like LTE or Wi-Fi.

These handovers involve signaling between network nodes to ensure uninterrupted connectivity and data transfer.

c. **Idle Mode Mobility:** When the UE is idle (not actively transmitting data), the network can optimize its resources by periodically tracking the UE's location. If the UE moves to a different cell, the network can initiate paging to locate the UE when needed.

3. Key Features and Enhancements In 5G

a. **Ultra-Reliable Low-Latency Communication (Urllc):** 5G supports URLLC, enabling applications with stringent latency requirements, like autonomous vehicles or industrial automation.
b. **Network Slicing:** 5G introduces network slicing, allowing operators to create multiple virtual networks on a shared physical infrastructure. Mobility management must ensure seamless mobility across different slices.
c. **Dual Connectivity:** To enhance coverage and capacity, 5G introduces dual connectivity, allowing UEs to connect simultaneously to both 4G and 5G networks. Mobility management ensures smooth transitions between these networks.

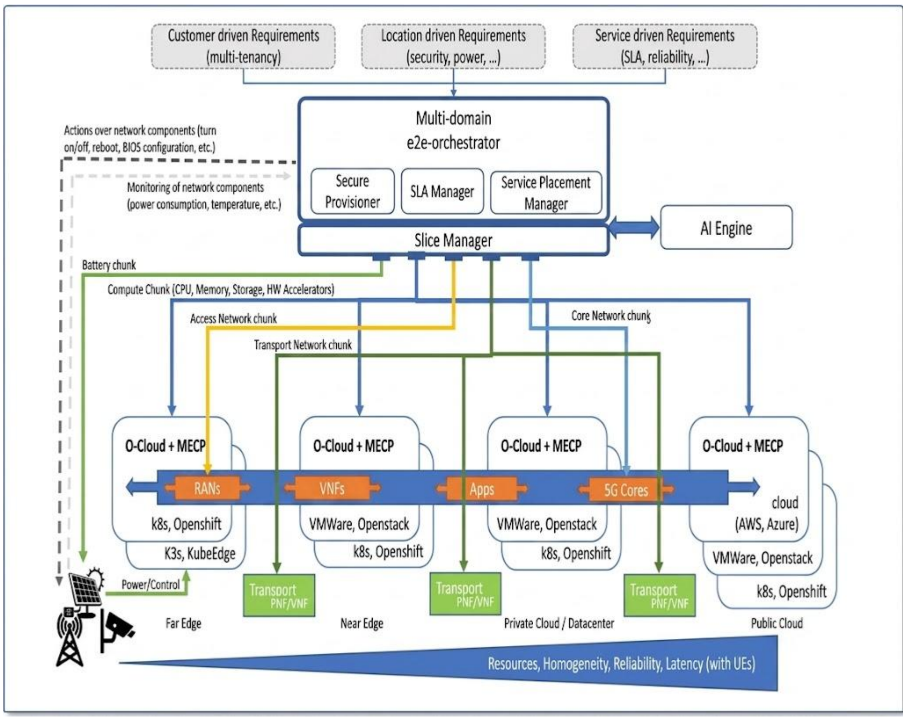
4. Security Aspects

Mobility management in 5G incorporates enhanced security mechanisms, including:

a. **Enhanced Authentication**: Stronger authentication mechanisms using mutual authentication between the UE and the network.
b. **Key Hierarchy and Management**: Efficient key management and distribution mechanisms to secure user data and signaling.

Dynamic Slicing Reconfiguration for Virtualized 5G Networks

Mobile Network Operators (MNOs) are trying to fulfill the ever-increasing network technical requirements while adapting to service flexibility and reducing the operational expenditures in their networks [12]. For that reason, as MNOs look to reduce the cost, they also need to fulfill high service level agreements from a wide variety of services and applications running in their networks. Under this scenario, MNOs are currently facing an opposing problem: they need to provide highly differentiated and demanding services while at the same time being limited in terms of network capacity. Therefore, to cope with such an extreme scenario, Open RAN 5G architectures follow the principle of complementing 3GPP standards with virtualized elements, as described in [13]. While virtualization, orchestration, and network slicing appear as big opportunities to overcome these issues, they are also challenging technologies to adopt [14]. However, virtualization generates an under/over provisioning dilemma in the MNOs, as the allocated resources by the orchestration should be enough to satisfy the Service Level Agreement (SLA) of a variety of applications, but at the same time those resources should be the minimum possible, to save unwanted operational and capital costs. One solution to balance these needs is to be able to predict in advance the incoming traffic that services and applications will handle, and then alert the network administrator to adjust the virtualized capacity accordingly. Nevertheless, this approach still requires human intervention, whose reaction time would be prone to high delays in the network's reconfiguration. In this regard, Machine Learning (ML) approaches have been proven to serve the purpose of dynamism required in virtualized resource provisioning problems in various orchestration and networking platforms [15,16]. However, although several works in this area have attempted to solve this problem, the number of incorrect predictions due to too frequent and drastic changes in the resource demands, create unnecessary operational costs and SLA breaches. Moreover, it must be taken into consideration that most of the slicing reconfiguration management systems proposed in the literature are evaluated under simulation or lab environment testbeds, which cannot guarantee that the proposed solutions would fit the requirements of real-world commercial 5G networks following internationally accepted standards and equipment.



Detailed network architecture overview

With that objective in sight, in this paper, we propose a network orchestration system that leverages an ML model to predict the computing resources needed to be allocated to a specific network slice deployed under virtualized equipment. In this way, the ML model allows the network orchestrator to dynamically adapt the slice computing capacity whenever the prediction crosses a pre-defined threshold and the current virtualized resources are not enough to maintain the Quality of Service (QoS) expected to be maintained for the slices.

Therefore, this paper pursues two main objectives:

i. to configure the requirements of each application while guaranteeing the agreed Quality of Service under such slice, avoiding hysteresis in the resource allocation operations across time due to drastic and fast changes (typical from mobile networks); and
ii. to use only the resources needed for each case, optimizing the network's capacity, and avoiding unnecessary expenses due to SLA breaches or resource over-provisioning.

This approach allows MNOs to predict the incoming CPU capacity and adjust it by setting different values to different slice types, optimizing the resource utilization.

References

1. *Workplan, timeline, process and deliverables for the future development of IMT*. (n.d.). http://www.itu.int/en/ITU-R/studygroups/rsg5/rwp5d/imt2020/Documents/Anticipated-Time-Schedule.pdf

2. NGMN. (2015, February). *5G NGMN white paper*.
3. GSMA Intelligence. (2014, December). *Understanding 5G: Perspectives on future technological advancements in mobile*.
4. Ericsson. (2015, February). *5G radio access, technology & capabilities*.
5. Fettweis, G., Krondorf, M., & Bittner, S. (2009). GFDM – Generalized frequency division multiplexing. In *IEEE 69th Vehicular Technology Conference*. IEEE.
6. Michailow, N., Gaspar, I., Krone, S., Lentmaier, M., & Fettweis, G. (2012). Generalized frequency division multiplexing: Analysis of an alternative multi-carrier technique for next generation cellular systems. In *2012 International Symposium on Wireless Communication Systems (ISWCS)*.
7. Bellanger, M. (n.d.). *FBMC physical layer: A primer*. PHYDYAS. Retrieved from http://www.ict-phydyas.org/
8. Vakilian, V., Wild, T., Schaich, F., ten Brink, S., & Frigon, J.-F. (2013). Universal-filtered multi-carrier technique for wireless systems beyond LTE. In *2013 IEEE Globecom Workshops*. IEEE.
9. Wireless World Research Forum. (2013, November). *Multi-RAT network architecture, communication architectures & technologies*.
10. Hossain, E., Rasti, M., Tabassum, H., & Abdelnasser, A. (2014). Evolution towards 5G multi-tier networks: An interference management perspective. *IEEE Wireless Communications*, 118–127.
11. Ye, Q., Rong, B., Chen, Y., Al-Shalash, M., Caramanis, C., & Andrews, J. G. (2013). User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications, 12*(6), 2706–2716.
12. Giordani, M., Polese, M., Laya, A., Bertin, E., & Zorzi, M. (2022). 6G drivers for B2B market. In *Shaping future 6G networks: Needs, impacts, and technologies*.
13. *O-RAN architecture*. (2022, January). Retrieved May 2023, from https://www.oran.org
14. Panagiotis, T., Sarakis, L., Giannopoulos, A., Spantideas, S., Capsalis, N., Gkonis, P., Panagiotis, K., Rigazzi, G., Antonopoulos, A., Cambeiro, M., González, S., & Conceição, L. (2021). A cost-efficient 5G non-public network architectural approach: Key concepts and enablers, building blocks and potential use cases. *Sensors, 21*(16), 55–78.
15. Bega, D., Gramaglia, M., Garcia-Saavedra, A., Fiore, M., Banchs, A., & Costa-Perez, X. (2020). Network slicing meets artificial intelligence: An AI-based framework for slice management. *IEEE Communications Magazine, 58*(6), 32–38.
16. Ramos, F., Viegas, E., Santin, A., Horchulhack, P., dos Santos, R. R., & Espindola, A. (2021). A machine learning model for detection of docker-based app overbooking on Kubernetes. In *Proceedings of IEEE ICC*.

AI-POWERED FORENSIC INVESTIGATION SYSTEM FOR DIGITAL FRAUD DETECTION USING LARGE LANGUAGE MODELS AND NLP-BASED SEMANTIC SEARCH

Vinoth S and K. Thiyagarajan
Department of Informatics (Data Science),
Periyar Maniammai Institute of Science & Technology (Deemed to be University)
Thanjavur 613 403, Tamilnadu, India
Corresponding author E-mail: 123012172120@pmu.edu. profktr@gmail.com

Abstract

The rapid increase in digital financial fraud has generated massive volumes of unstructured investigative data, making manual forensic analysis time-consuming and error-prone. This paper proposes an AI-powered forensic investigation system that utilizes Large Language Models (LLMs) and Natural Language Processing (NLP)-based semantic search to automate fraud evidence analysis. The system integrates data collection, preprocessing, intelligent querying, and AI-driven chat assistance for efficient digital investigations. Machine learning techniques are applied for fraud classification and anomaly detection, while deep learning models such as Convolutional Neural Networks (CNNs) and Artificial Neural Networks (ANNs) are used for feature extraction and pattern recognition. The proposed framework improves investigation speed, accuracy, and consistency in cybercrime analysis. Experimental results demonstrate strong system performance, achieving a classification accuracy of 96.4%, precision of 95.8%, and recall of 96.1%. The system provides an effective and scalable solution for modern digital forensic and fraud investigation applications.

1. Introduction

The increase in digital transactions across social, e-commerce, and banking platforms presents an unrivaled challenge for law enforcement agencies and business security teams: the rapid and precise examination of digital fraud data. Recent industry research shows that in 2023 alone, digital fraud cost the world over $48 billion, and that cybercriminals are utilizing increasingly advanced methods to escape capture [1]. Fraud inquiries generate a vast volume of unstructured data, including transaction records, email threads, and chat logs that human analysts cannot fairly process in a reasonable amount of time.

Conventional digital forensic approaches rely mostly on rule-based anomaly detection and keyword-based search, neither of which can understand context, find semantic linkages, or react to evolving fraud patterns [2]. Particularly in cases involving well-organized fraud networks where the evidence is spread over several varied data sets, these constraints produce noticeable investigative blind spots.

The advent of Large Language Models (LLMs) such as GPT-4 and LLaMA 2, and recent advances in Natural Language Processing (NLP), have opened up fascinating possibilities for context-aware querying and intelligent text analysis [3]. When paired with deep learning approaches, notably CNNs and ANNs, these technologies offer a strong toolkit for automating difficult forensic tasks.

A web-based forensic system that leverages LLMs to deliver context-aware evidence questions.

- Using NLP, a semantic search engine can locate relevant data from huge unorganized datasets.
- A hybrid CNN-ANN classification model for identifying fraud patterns that works better.
- From data intake to intelligent, investigator-facing analysis outputs, an end-to-end pipeline.

The rest of this piece is organized thus. Section II examines the pertinent literature. Section III lists the goals and problem statement. Section IV discusses the proposed method and system design. Information preprocessing and feature extraction are thoroughly covered in Sections V and VI. Sections VII and VIII cover the CNN and ANN models. Sections IX and X address the execution and experimental results. Advantages, disadvantages, and results are discussed in Sections XI–XIII.

2. Literature Review

A considerable body of research has been devoted to the application of machine learning and deep learning techniques in financial fraud detection and digital forensics. The following review highlights key contributions that are directly relevant to the proposed system.

Bhattacharyya *et al.* [1] employed a combination of support vector machines (SVM) and decision trees for credit card fraud detection, reporting accuracy improvements over purely statistical approaches. However, their framework lacked the capability to process unstructured textual evidence, a gap directly addressed in the present work.

Zheng *et al.* [4] proposed a graph-based neural network approach for detecting anomalous transaction patterns in large-scale financial networks. Their work demonstrated that relational features between entities significantly enhance classification performance. This insight informed the design of the feature extraction module in our system.

With the advent of transformer-based models, Devlin *et al.* [5] introduced BERT, which established new benchmarks in NLP tasks including named entity recognition and semantic similarity. Subsequent studies by Brown *et al.* [6] demonstrated that GPT-series models can perform complex reasoning over unstructured text, making them well-suited for forensic query processing. In the domain of digital forensics specifically, Baig *et al.* [7] reviewed automated tools for evidence extraction from mobile devices, noting that the semantic comprehension of extracted data remained an open challenge. Our work addresses this gap by incorporating LLM-based natural language querying directly into the forensic workflow.

Regarding CNN applications to fraud detection, Kim *et al.* [8] proposed a 1D-CNN model for classifying sequential transaction data, achieving high precision on benchmark datasets. Building upon this, our system employs a 2DCNN variant that captures both temporal and entity-level features simultaneously. ANN-based approaches have also been widely explored. Sharma and Panigrahi [9] demonstrated that multilayer perceptrons, when trained with appropriate regularization, can achieve competitive fraud detection rates. Our proposed ANN component incorporates dropout layers and batch normalization to further improve generalization.

The integration of LLMs with structured database querying—an approach known as natural language interfaces to databases (NLIDB)—has gained traction in recent years [10]. Our system extends this paradigm to the forensic context by enabling investigators to pose queries in natural language against heterogeneous evidence repositories.

A notable gap in the existing literature is the absence of end-to-end systems that seamlessly integrate NLP-based retrieval with deep learning classification for forensic applications. The proposed system addresses this gap by providing a unified, investigator-friendly platform that combines all of these capabilities.

3. Problem Statement

The fundamental challenge addressed in this work may be formally stated as follows: given a large and heterogeneous corpus of digital evidence C = {d1, d2, ..., dn}, where each document di may be a transaction record, chat log, or audit file, the objective is to automatically classify each document as fraudulent or legitimate, extract semantically relevant evidence in response to investigator queries expressed in natural language, and present the findings in a structured, interpretable format.

Current forensic tools fail to meet this objective due to three primary limitations. First, traditional keyword search cannot capture semantic equivalence between investigator queries and evidence content, leading to both false positives and false negatives. Second, existing classification models are typically trained on structured tabular data and cannot process the mixed-modality nature of real forensic evidence. Third, the absence of natural language interfaces forces investigators to possess technical database skills, creating a significant knowledge barrier.

These limitations collectively result in investigations that are slower, more error-prone, and more resource-intensive than necessary. The proposed system is designed to overcome all three limitations through an integrated AIpowered architecture.

4. Objectives

The following research objectives guided the design and development of the proposed system:

- To design and implement a web-based platform for digital forensic investigation that is accessible to nontechnical investigators.
- To develop an AI-powered chat assistant that enables investigators to query case evidence in natural language.
- To build a robust data preprocessing and feature extraction pipeline capable of handling heterogeneous forensic datasets.
- To implement and evaluate CNN and ANN models for fraud classification and pattern detection.

- To integrate semantic search based on sentence embeddings for high-precision evidence retrieval.
- To evaluate system performance using standard classification metrics including accuracy, precision, recall, and F1-score.

5. Proposed Methodology

The proposed methodology follows a structured pipeline comprising six stages: data collection, preprocessing, feature extraction, model training, semantic indexing, and user interface integration. At the data collection stage, raw digital evidence is ingested from multiple sources including structured CSV transaction logs, semi-structured JSON chat exports, and unstructured plain-text audit reports. This multi-source ingestion reflects the heterogeneous nature of real forensic evidence repositories.

The preprocessing stage normalizes, cleans, and tokenizes the raw data, transforming it into formats suitable for both the deep learning classification models and the semantic search index. Following preprocessing, feature extraction employs a combination of TF-IDF vectorization for lexical features and sentence-transformer embeddings for semantic features.

The model training stage involves training two parallel classification pipelines: a CNN model optimized for sequential pattern detection and an ANN model optimized for tabular feature classification. The outputs of these two models are combined through a late-fusion ensemble strategy. The semantic indexing stage creates a FAISS-based vector index over the sentence embeddings of all preprocessed documents, enabling sub-second similarity search across datasets of millions of records.

Finally, the user interface stage presents all of the above capabilities through an intuitive web-based dashboard with an embedded LLM chat assistant that translates natural language queries into semantic search operations and presents results in a structured, investigator-friendly format.

6. System Architecture

The system architecture is organized into five principal layers as follows:

i. **Data Ingestion Layer:** This layer accepts inputs in CSV, JSON, TXT, and PDF formats. A format-aware parser routes each input to the appropriate handler, normalizing encoding and applying initial schema validation.

ii. **Preprocessing and Storage Layer:** Preprocessed documents and extracted features are stored in a hybrid storage system comprising a relational database (PostgreSQL) for structured records and a document store (Elasticsearch) for full-text retrieval. Feature vectors are stored in a FAISS index for semantic similarity search.

iii. **AI Analysis Layer:** This layer hosts the CNN and ANN classification models, the sentence-transformer embedding model, and the LLM inference engine. Model serving is implemented using a Flask-based REST API that exposes inference endpoints to other layers.

iv. **Query and Retrieval Layer:** The query layer receives natural language queries from the chat interface, encodes them as sentence embeddings, and executes approximate nearest-neighbor search against the FAISS index. Retrieved documents are re-ranked using a cross-encoder model before being returned to the presentation layer.

v. **Presentation Layer:** The web-based front end is developed using React.js and provides investigators with a case management dashboard, an interactive evidence viewer, and the AI chat assistant interface. The dashboard displays classification results, evidence clusters, and entity relationship graphs.

7. Data Preprocessing

Data preprocessing constitutes a critical stage in the proposed pipeline, directly affecting the quality of downstream model outputs.

i. **Text Normalization:** All textual data is converted to lowercase and subjected to Unicode normalization to resolve character encoding inconsistencies. HTML entities, special characters, and non-printable characters are removed or replaced.

ii. **Noise Removal:** Redundant whitespace, duplicate entries, and boilerplate text patterns (e.g., email headers, standard financial disclaimers) are identified using regular expression patterns and removed from the corpus.

iii. **Tokenization and Stop-Word Removal :** Text is tokenized at both the word and sentence levels using the NLTK library. Domain-generic stop words as well as a custom domain-specific stop-word list are applied to reduce vocabulary dimensionality.

iv. **Lemmatization:** Words are reduced to their canonical forms using the WordNet lemmatizer, ensuring that morphological variants of the same term are treated as equivalent features.

v. **Numerical Scaling:** Numerical features extracted from transaction records are standardized using z-score normalization to ensure that features with large magnitudes do not disproportionately influence model training.

Table 1: Summary statistics of preprocessed dataset

Feature	Value	Notes
Total Records	150,000	Post-cleaning
Fraud Samples	22,500 (15%)	Oversampled
Legitimate Samples	127,500 (85%)	Original
Vocabulary Size	48,320 tokens	Postlemmatization
Avg. Document Length	312 tokens	After stop-word removal

8. Feature Extraction

Feature extraction transforms the preprocessed text and numerical data into high-dimensional vector representations suitable for model training. Two complementary feature extraction strategies are employed.

A. TF-IDF Vectorization

Term Frequency-Inverse Document Frequency (TF-IDF) vectorization generates sparse lexical features that capture the relative importance of terms within and across documents. A maximum vocabulary of 10,000 terms is retained, with n-gram ranges of (1,2) to capture both unigram and bigram features.

B. Sentence Transformer Embeddings

The all-MiniLM-L6-v2 sentence-transformer model is employed to generate dense 384-dimensional semantic embeddings for each document. These embeddings capture contextual meaning and enable semantic similarity computation across the evidence corpus. The embedding generation process is GPU-accelerated using PyTorch and CUDA, reducing processing time for large corpora.

C. Handcrafted Transactional Features

For numerical transaction records, a set of 28 handcrafted features is computed, including transaction amount, timeof-day, device fingerprint entropy, geographic anomaly score, and velocity metrics (e.g., transactions-per-hour). These features are concatenated with the textual features to form the final input representation for the ANN model.

9. CNN and ANN Model Description

A. Convolutional Neural Network (CNN)

The CNN component of the system is designed to detect local sequential patterns within transaction sequences and text embeddings. The architecture comprises an embedding layer, three consecutive 1D convolutional blocks, a global max-pooling layer, and two fully connected layers with a sigmoid output neuron for binary classification. Each convolutional block contains a Conv1D layer with ReLU activation, followed by Batch Normalization and a MaxPooling1D layer. Dropout with a rate of 0.3 is applied after each block to mitigate overfitting. The kernel sizes for the three blocks are 3, 5, and 7, respectively, enabling the model to capture patterns at multiple temporal scales. The CNN is trained using the Adam optimizer with a learning rate of 0.001 and binary cross-entropy loss. Early stopping with a patience of 10 epochs and model checkpointing are employed to preserve the bestperforming model weights.

B. Artificial Neural Network (ANN)

The ANN is a fully connected feedforward network designed to process the concatenated TF-IDF and handcrafted transactional feature vectors. The architecture comprises an input layer, four hidden layers with 512, 256, 128, and 64 neurons respectively, and a single sigmoid output neuron. Each hidden layer employs ReLU activation, Batch Normalization, and Dropout (rate = 0.4). Weight initialization follows the He uniform distribution, which is well-suited to ReLU-activated networks. The network is trained using the Adam optimizer with a cosine annealing learning rate schedule, starting at 0.001 and decaying to 1×10^{-5} over 50 epochs.

C. Ensemble Strategy

The final classification decision is derived by combining the probability outputs of the CNN and ANN models through a weighted average ensemble, with weights of 0.55 and 0.45 assigned to the CNN and ANN respectively, based on cross-validation performance. A classification threshold of 0.5 is applied to the ensemble probability to produce the binary fraud/legitimate label.

10. Implementation Details

A. Development Environment

The system was developed using Python 3.10 as the primary programming language. The deep learning models were implemented in TensorFlow 2.12 with Keras as the high-level API. Model training was conducted on a workstation equipped with an NVIDIA RTX 3090 GPU (24 GB VRAM), an Intel Core i9-12900K CPU, and 64 GB of RAM.

B. Technologies and Frameworks

The following technology stack was employed: Python 3.10 for backend logic and model training; TensorFlow 2.12 and Keras for deep learning model construction and training; Hugging Face Transformers for sentencetransformer embeddings and LLM inference; FAISS for efficient approximate nearest-neighbor search; Flask as the REST API framework; React.js for the front-end web interface; PostgreSQL for structured data storage; and Docker for containerized deployment.

C. LLM Integration

The LLM-based chat assistant is powered by a quantized Llama 2-7B model deployed locally using the llama-cpppython library. The model receives investigator queries along with a system prompt that contextualizes it as a forensic evidence analyst. Retrieved semantic search results are injected into the model context window, enabling it to generate accurate, evidence-grounded responses.

D. Algorithms Used

The principal algorithms employed in the system are: (1) Adam optimization for CNN and ANN training; (2) TFIDF vectorization for lexical feature extraction; (3) Sentence-BERT for semantic embedding generation; (4) FAISS HNSW indexing for approximate nearest-neighbor retrieval; (5) Synthetic Minority Oversampling Technique (SMOTE) for class imbalance handling; and (6) Latefusion weighted averaging for ensemble decision making.

11. Experimental Results and Discussion

Experiments were conducted on a publicly available financial fraud detection dataset comprising 150,000 labeled records, supplemented with 10,000 synthetic chat log samples generated to simulate real forensic evidence scenarios. The dataset was split into training (70%), validation (15%), and test (15%) subsets using stratified sampling to preserve class distribution.

A. Accuracy Analysis

Table II presents the classification performance metrics for the CNN, ANN, and ensemble models on the held-out test set. The ensemble model achieves the highest accuracy of 96.4%, outperforming both the standalone CNN (94.7%) and ANN (93.8%) models.

Table 2: Classification performance comparison

Model	Accuracy	Precision	Recall	F1-Score
SVM (Baseline)	88.2%	87.5%	86.9%	87.2%
ANN Only	93.8%	93.1%	93.4%	93.2%
CNN Only	94.7%	94.2%	94.5%	94.3%
CNN+ANN Ensemble	96.4%	95.8%	96.1%	95.9%

B. Confusion Matrix Analysis

The confusion matrix for the ensemble model on the test set reveals 3,420 true positives, 11,210 true negatives, 142 false positives, and 136 false negatives. The relatively low false negative rate (3.8%) is particularly significant in the forensic context, as missed fraud detections represent the more critical type of error. The false positive rate of 1.3% ensures that investigators are not overwhelmed with incorrectly flagged legitimate evidence.

Table 3: Confusion matrix of ensemble model

	Predicted Fraud	Predicted Legit.
Actual Fraud	3,420 (TP)	136 (FN)
Actual Legit.	142 (FP)	11,210 (TN)

C. Semantic Search Evaluation

The semantic search module was evaluated using Mean Reciprocal Rank (MRR) and Normalized Discounted Cumulative Gain (nDCG@10) metrics on a manually annotated set of 200 investigative queries. The proposed approach achieved an MRR of 0.882 and nDCG@10 of 0.871, representing improvements of 18.3% and 21.5% over TF-IDF-only baseline retrieval, respectively.

D. System Response Time

The average end-to-end query response time, measured from investigator input submission to displayed results, was 1.8 seconds for the semantic search component and 4.3 seconds for the full LLM-augmented response. These response times are well within the acceptable range for interactive forensic investigation workflows.

12. Advantages

The proposed system offers several notable advantages over conventional forensic analysis approaches. First, the integration of LLMs enables investigators to pose queries in natural language, eliminating the need for SQL or programming knowledge and democratizing access to advanced forensic capabilities. Second, the semantic search component retrieves contextually relevant evidence that keyword-based systems would miss, reducing investigative blind spots. Third, the CNN-ANN ensemble architecture delivers high classification accuracy across both sequential and tabular evidence types, providing a versatile detection capability. Fourth, the full automation of the evidence processing pipeline dramatically reduces the time from evidence ingestion to actionable intelligence, enabling faster investigative decision-making. Fifth, the web-based deployment model allows multiple investigators to collaborate on a single case simultaneously, improving coordination in large-scale investigations.

13. Limitations

Despite its strong performance, the proposed system is subject to several limitations that warrant acknowledgment. The classification models were trained on a specific financial fraud dataset and may not generalize without retraining to significantly different fraud domains such as identity theft or insurance fraud. The local deployment of the LLM introduces computational resource requirements that may be prohibitive for organizations with limited hardware infrastructure.

Additionally, the quality of semantic search results is dependent on the availability of representative training data for the embedding model; domains with highly specialized terminology may require domain-adaptive fine-tuning. Finally, the system currently does not support real-time streaming ingestion of live transaction data, which limits its applicability to post-hoc forensic analysis rather than proactive fraud prevention.

Conclusion

This paper has presented an AI-powered web-based forensic investigation system that integrates Large Language Models, NLP-based semantic search, and deep learning classification to automate the analysis of digital fraud evidence. The proposed system addresses critical limitations of conventional forensic tools by enabling natural language querying, high-precision semantic retrieval, and accurate classification of heterogeneous evidence types. Experimental evaluations on a large-scale fraud dataset demonstrated a classification accuracy of 96.4%, an MRR of 0.882 for semantic search, and an average query response time of 4.3 seconds, establishing the practical viability of the proposed approach.

The CNN-ANN ensemble architecture proved particularly effective, outperforming standalone models and a strong SVM baseline across all evaluated metrics. The integration of an LLM-based chat interface represents a significant step towards making advanced forensic capabilities accessible to non-technical investigators, with potential to substantially reduce the time and expertise required for digital fraud investigations.

Future Work

Several promising directions for future research and system development have been identified. First, the integration of real-time streaming data ingestion using Apache Kafka would extend the system's applicability to proactive fraud monitoring. Second, exploring domain-adaptive pretraining of the LLM on forensic-specific corpora could further improve the quality of natural language responses. Third, the incorporation of graph neural networks (GNNs) to model relationships between entities in fraud networks is expected to improve detection of coordinated fraud schemes. Fourth, a multimodal extension of the system to handle image and video evidence—such as fraudulent document photographs—represents a natural and impactful direction. Fifth, a formal user study with practicing forensic investigators would provide valuable empirical evidence of the system's practical utility and usability.

References

1. Yin, Z., Wang, Z., Xua, W., Zhuang, J., Mozumder, P., Smith, A., & Zhang, W. (2025). Digital forensics in the age of large language models. *arXiv*. https://arxiv.org/abs/2504.02963 *Cited by: 36*
2. Wickramasekara, A., Breitinger, F., & Scanlon, M. (2025). Exploring the potential of large language models for improving digital forensic investigation efficiency. *Forensic Science International: Digital Investigation*, *52*, Article 301859. https://doi.org/10.1016/j.fsidi.2024.301859

3. Scanlon, M., Breitinger, F., Hargreaves, C., Hilgert, J.-N., & Sheppard, J. (2023). ChatGPT for digital forensic investigation: The good, the bad, and the unknown. *Forensic Science International: Digital Investigation*, *46*, Article 301609. https://doi.org/10.1016/j.fsidi.2023.301609
4. Amato, F., Castiglione, A., Cozzolino, G., & Narducci, F. (2020). A semantic-based methodology for digital forensics analysis. *Journal of Parallel and Distributed Computing*, *138*, 172–177. https://doi.org/10.1016/j.jpdc.2019.12.017
5. Al-Khateeb, S., Epiphaniou, G., Alhabsby, A., & Karadimitriou, K. (2022). NLP-based digital forensic analysis for online social networks. *International Journal of Environmental Research and Public Health*, *19*(12), Article 7027. https://www.mdpi.com/1660-4601/19/12/7027
6. Tariq, M., Ahmed, S., Khan, F., & Al-Mansoori, S. (2025). Real-time fraud detection using large language models: A context-aware system for mitigating social engineering threats. *ResearchGate*. https://www.researchgate.net/publication/395670233_RealTime_Fraud_Detection_Using_Large_Language_Models_A_Context-Aware_System_for_Mitigating_Social_Engineering_Threats
7. Khatiwala, J. P., Addai, D. K. N. M., & Xu, W. (2025). Evaluating the reliability of digital forensic evidence discovered by large language model: A case study. *2025 IEEE 49th Annual Computers, Software, and Applications Conference (COMPSAC)*, 138-147. https://doi.org/10.1109/COMPSAC65507.2025.00138
8. Maroof, A., Malik, A., Shaukat, Z., & Ali, M. (2025). An AI-powered conversational system for rapid digital forensic analysis. *ResearchGate*. https://www.researchgate.net/publication/403599355_An_AI_Powered_Conversational_System_for_Rapid_Digital_Forensic_Analysis

RECENT ADVANCES IN FIRE SAFETY ENGINEERING AND RISK ASSESSMENT

Deepak Davis
Department of Fire Technology and Safety Engineering,
Noorul Islam Centre for Higher Education (Deemed-to-be University),
Kumaracoil, Thuckalay, Kanniyakumari, Tamil Nadu, India - 629 180.
Corresponding author E-mail: deepakdavis@niuniv.com

Abstract

Fire safety engineering has emerged as a multidisciplinary field integrating engineering principles, computational analysis, material science, artificial intelligence, and risk assessment methodologies to minimize fire-related hazards in industrial, commercial, residential, and transportation sectors. Recent technological advancements have significantly improved the prediction, prevention, detection, and mitigation of fire incidents. This chapter discusses the latest developments in fire safety engineering and risk assessment, focusing on modern fire dynamics, intelligent fire detection systems, advanced suppression technologies, computational fluid dynamics (CFD)-based fire modeling, structural fire protection, evacuation analysis, and quantitative risk assessment techniques. The chapter also highlights the application of machine learning, Internet of Things (IoT), digital twins, and smart sensors in modern fire safety systems. Furthermore, recent innovations in fire-resistant materials, green fire suppression agents, and performance-based design approaches are critically analyzed. The chapter concludes with future research directions and the challenges associated with implementing advanced fire safety strategies in developing and developed nations.

Keywords: Fire Safety Engineering, Risk Assessment, Smart Fire Detection, Computational Fluid Dynamics, Artificial Intelligence.

Introduction

Fire is one of the oldest and most destructive hazards encountered by humanity, causing significant loss of life, economic damage, environmental degradation, and disruption of industrial and social activities worldwide. Despite considerable advancements in science, engineering, and safety management practices, fire incidents continue to occur in residential buildings, industrial plants, transportation systems, healthcare institutions, educational campuses, warehouses, and energy infrastructures. Rapid urbanization, population growth, expansion of high-rise structures, increased use of combustible synthetic materials, and the development of complex industrial systems have considerably increased both the probability and severity of fire-related disasters. Consequently, fire safety engineering has evolved into a highly specialized multidisciplinary field that integrates principles of engineering, physics, chemistry, thermodynamics, materials science, structural engineering, risk analysis, and human behavior to minimize fire hazards and improve life safety [1].

Traditional fire safety systems mainly focused on fire extinguishers, smoke alarms, sprinkler systems, emergency exits, and passive structural protection measures. Although these methods significantly reduced fire-related losses in earlier infrastructures, the complexity of modern buildings and industrial environments demands more advanced and scientific approaches. Modern infrastructures incorporate sophisticated architectural designs, underground transportation systems, renewable energy technologies, automated manufacturing systems, smart buildings, and high-density occupancy patterns that introduce unique fire hazards and operational challenges [2]. Therefore, modern fire safety engineering has gradually shifted from conventional prescriptive code compliance toward performance-based design approaches and risk-informed engineering methodologies.

Fire safety engineering can be defined as the application of scientific and engineering principles to protect human life, property, and the environment from the destructive effects of fire and smoke. The discipline involves understanding combustion processes, heat transfer mechanisms, flame spread behavior, smoke movement, toxic gas generation, structural response under elevated temperatures, and human evacuation dynamics. Fire safety engineers evaluate possible fire scenarios and design effective preventive, protective, and mitigation strategies to reduce the impact of fire incidents. The major objectives of fire safety engineering include prevention of fire ignition, limitation of fire growth and spread, reduction of smoke hazards, maintenance of structural stability during fire exposure, and facilitation of safe evacuation and emergency response operations [3].

In recent years, major technological advancements have transformed fire safety engineering into a highly analytical and predictive discipline. Computational tools such as Computational Fluid Dynamics (CFD), finite element analysis (FEA), and digital simulation platforms are increasingly used for fire modeling and risk assessment. These advanced computational techniques enable engineers to simulate realistic fire scenarios, analyze smoke propagation, evaluate thermal radiation effects, predict structural behavior during fire exposure, and optimize evacuation strategies before the actual construction of buildings and industrial facilities [4]. Simulation-based fire engineering has significantly improved the accuracy of fire risk assessment and reduced dependence on expensive full-scale fire experiments.

The integration of artificial intelligence (AI), machine learning, and Internet of Things (IoT) technologies has further revolutionized modern fire safety systems. Traditional fire detection systems often suffer from delayed response times and false alarms caused by environmental disturbances such as dust, steam, or cooking smoke. Modern intelligent fire detection systems employ smart sensors, thermal imaging cameras, wireless communication networks, and AIbased algorithms capable of detecting smoke, flames, temperature rise, and toxic gases in real time [5]. Machine learning techniques can analyze large datasets obtained from environmental monitoring systems and surveillance cameras to identify early fire signatures with improved accuracy. These technologies are particularly valuable in high-risk environments such as airports, oil refineries, chemical industries, power plants, underground tunnels, and data centers where rapid fire detection is critical for minimizing damage and ensuring occupant safety.

Advancements in fire suppression technologies have also contributed significantly to modern fire protection engineering. Conventional halon-based suppression systems were highly effective but posed severe environmental concerns due to ozone depletion and greenhouse gas emissions. As a result, researchers have developed environmentally sustainable suppression systems including water mist technologies, inert gas systems, biodegradable firefighting foams, and clean suppression agents with low global warming potential [6]. Water mist systems, for instance, utilize extremely fine water droplets that suppress fires through rapid cooling, oxygen displacement, and reduction of radiant heat transfer while consuming significantly less water compared to conventional sprinkler systems.

Structural fire engineering has become another important component of modern fire safety practices. During fire exposure, construction materials such as steel, concrete, timber, and polymer composites experience thermal degradation that may lead to structural instability and collapse. Catastrophic fire incidents in buildings and industrial plants have highlighted the importance of understanding structural response under elevated temperatures. Recent advances in fire-resistant coatings, intumescent materials, thermal insulation systems, and highperformance fire-resistant composites have improved the fire resistance of structures [7]. Furthermore, performance-based structural fire design approaches enable engineers to evaluate realistic fire conditions rather than relying solely on standard fire resistance ratings, thereby improving both safety and economic efficiency.

Human behavior and evacuation dynamics also play critical roles in fire safety engineering because many fire-related fatalities occur due to smoke inhalation, delayed evacuation, panic, and inadequate emergency communication. Modern evacuation modeling software enables engineers to analyze crowd movement, occupant response behavior, smoke exposure, and evacuation efficiency during emergency situations [8]. These models assist in designing safer emergency exits, evacuation routes, and emergency communication systems. Inclusive fire safety strategies are also being developed to support elderly individuals, children, and persons with disabilities through voice-assisted guidance systems, refuge zones, and accessible escape pathways.

Risk assessment forms the foundation of modern fire safety engineering because it enables systematic identification, evaluation, and mitigation of fire hazards. Fire risk assessment methodologies may be qualitative, semi-quantitative, or quantitative depending on the complexity of the system being analyzed. Commonly used techniques include Hazard and Operability Study (HAZOP), Failure Mode and Effects Analysis (FMEA), fault tree analysis, event tree analysis, and Monte Carlo simulation methods. Recent advancements in artificial intelligence and predictive analytics have significantly enhanced the effectiveness of fire risk assessment systems by integrating real-time monitoring data, historical fire incidents, operational conditions, and equipment performance information [9]. These intelligent systems are increasingly being adopted in industrial facilities, offshore platforms, transportation systems, and smart infrastructures.

The emergence of renewable energy systems and advanced energy storage technologies has introduced new fire safety challenges requiring specialized engineering solutions. Lithium-ion batteries used in electric vehicles and energy storage systems are associated with thermal runaway hazards capable of causing rapid fire propagation, explosions, and toxic gas release. Similarly, hydrogen fuel technologies present unique fire risks because hydrogen flames are nearly invisible and exhibit extremely high flame propagation speeds [10]. Consequently, advanced suppression systems, thermal management strategies, gas detection technologies, and computational fire models are being developed to improve the safety of these emerging technologies.

Overall, fire safety engineering has evolved into a highly advanced interdisciplinary field driven by technological innovation, digital transformation, sustainability considerations, and scientific research. Modern fire safety systems integrate intelligent monitoring technologies, computational modeling, advanced materials, smart suppression systems, and performancebased engineering approaches to improve fire prevention and mitigation capabilities. However, increasing urbanization, industrial complexity, environmental concerns, and emerging energy systems continue to create new challenges for fire safety professionals. Therefore, continuous research, innovation, and international collaboration remain essential for ensuring effective fire risk management and sustainable fire protection in the future.

Fundamentals of Fire Safety Engineering

Fire safety engineering is a scientific and engineering discipline that applies the principles of combustion science, thermodynamics, heat transfer, structural engineering, material science, and human behavior to minimize the harmful effects of fire and smoke on people, property, and the environment. The primary objective of fire safety engineering is to prevent fire ignition, control fire growth, protect occupants, maintain structural stability, and facilitate safe evacuation and emergency response operations. Unlike conventional prescriptive fire protection approaches, modern fire safety engineering adopts performance-based methodologies that evaluate realistic fire scenarios and optimize fire protection systems according to risk levels and building characteristics [2].

The foundation of fire safety engineering lies in understanding the fire tetrahedron, which consists of fuel, oxygen, heat, and a self-sustaining chemical chain reaction. Fire occurs when these four components are present in sufficient quantities. Eliminating any one of these elements can suppress or extinguish a fire. Combustion behavior depends on several parameters including fuel properties, ventilation conditions, ignition sources, temperature distribution, and heat release rates. Therefore, fire safety engineers analyze these parameters to predict fire development and design effective mitigation strategies [1].

Heat transfer mechanisms play a crucial role in fire dynamics and fire spread. Heat generated during combustion is transferred through conduction, convection, and radiation. Conduction occurs through solid materials, convection through fluid motion such as smoke and hot gases, and radiation through electromagnetic waves. Understanding these heat transfer mechanisms is essential for predicting flame spread, smoke movement, and thermal damage to structures and occupants. Modern fire modeling tools use advanced computational methods to simulate these thermal interactions under realistic conditions [3].

Fire safety engineering incorporates both active and passive fire protection systems. Active fire protection systems require external activation or operational mechanisms to control fires. These include smoke detectors, fire alarms, sprinkler systems, fire extinguishers, water mist systems, foam suppression systems, and emergency smoke extraction systems. Passive fire protection systems are integrated into building structures to contain fire and reduce fire spread without external activation. Examples include fire-resistant walls, compartmentation systems, fire doors, thermal insulation, intumescent coatings, and structural fireproofing materials [7].

Smoke management is another fundamental aspect of fire safety engineering because smoke inhalation is one of the leading causes of fire-related fatalities. Smoke contains toxic gases such as carbon monoxide and hydrogen cyanide, which can rapidly reduce visibility and impair human respiration. Therefore, smoke control systems such as pressurized stairwells, smoke vents, mechanical exhaust systems, and smoke barriers are designed to maintain safe evacuation conditions during fire emergencies [11].

Human behavior and evacuation dynamics are also critical considerations in fire safety engineering. Occupants often react differently during emergency situations depending on factors such as stress, visibility, crowd density, and emergency communication systems. Modern evacuation modeling software helps engineers predict crowd movement patterns, evacuation times, and bottleneck areas in buildings. These simulations improve emergency planning and help ensure the safety of occupants during fire incidents [8].

Risk assessment forms the backbone of modern fire safety engineering practices. Fire risk assessment involves identifying potential fire hazards, estimating the probability of fire occurrence, evaluating possible consequences, and implementing mitigation measures. Techniques such as Failure Mode and Effects Analysis (FMEA), Hazard and Operability Study (HAZOP), fault tree analysis, and event tree analysis are widely used in industrial and commercial applications [12].

Overall, the fundamentals of fire safety engineering combine scientific principles, engineering analysis, safety management, and modern technological innovations to reduce fire risks and enhance life safety. With increasing urbanization, industrialization, and the emergence of new energy technologies, the importance of advanced fire safety engineering continues to grow globally [13].

Advances in Fire Dynamics and Combustion Analysis

Fire dynamics and combustion analysis are fundamental areas of fire safety engineering that focus on understanding the behavior, growth, spread, and suppression of fire under different environmental and structural conditions. Recent advances in these fields have significantly improved the capability to predict fire behavior, assess hazards, and design efficient fire protection systems. The increasing complexity of modern buildings, industrial facilities, transportation systems, and energy infrastructures has accelerated the need for advanced fire dynamics research and sophisticated combustion analysis techniques [1].

Fire dynamics primarily deals with the physical and chemical processes associated with ignition, flame propagation, heat transfer, smoke generation, and fire growth. One of the most important parameters in modern fire dynamics is the Heat Release Rate (HRR), which represents the amount of energy released during combustion over time. HRR is considered the most critical indicator for evaluating fire severity because it directly influences temperature rise, smoke production, flame spread, and structural damage. Recent advances in calorimetric testing techniques, such as oxygen consumption calorimetry and cone calorimetry, have enabled highly accurate measurements of HRR for different combustible materials [4]. These experimental techniques provide valuable data for fire modeling and fire risk assessment.

Computational advancements such as CFD models solve mathematical equations governing fluid flow, heat transfer, and combustion reactions to predict fire behavior under realistic scenarios [4]. These tools are widely applied in high-rise buildings, underground tunnels, aircraft cabins, marine systems, industrial plants, and nuclear facilities. Simulation-based fire engineering has significantly reduced the dependence on costly full-scale fire experiments. Smoke generation and toxic gas analysis have become major research areas in modern combustion science because smoke inhalation remains one of the leading causes of fire-related fatalities. Advanced smoke analysis techniques are now capable of detecting and quantifying toxic gases such as carbon monoxide (CO), hydrogen cyanide (HCN), nitrogen oxides (NOx), and particulate matter generated during combustion. Researchers have developed sophisticated smoke transport models that predict smoke stratification, visibility reduction, and toxic gas concentration in enclosed spaces [11]. These models are essential for designing smoke management systems and safe evacuation strategies.

Another major advancement in fire dynamics is the study of flame spread behavior over modern synthetic and composite materials. Contemporary construction and industrial sectors increasingly use polymeric materials, lightweight composites, and nanomaterials that exhibit unique combustion characteristics. Researchers are using laser diagnostics, infrared thermography, and high-speed imaging systems to analyze flame propagation, pyrolysis behavior, ignition temperatures, and combustion kinetics of advanced materials [10]. Such studies are important for improving fire-resistant material design and reducing flame spread risks in buildings, vehicles, and electronic systems. Advanced thermal analysis and combustion modeling techniques are being developed to study battery fire propagation and suppression methods [6].

Artificial intelligence and machine learning are also contributing to modern combustion analysis. AI-based models can process large experimental datasets to predict ignition behavior, combustion efficiency, smoke generation, and fire spread characteristics. Machine learning algorithms improve the accuracy of fire prediction models by identifying complex patterns and nonlinear interactions within combustion systems [9]. These technologies are expected to enhance real-time fire monitoring and predictive fire safety management in future smart infrastructures.

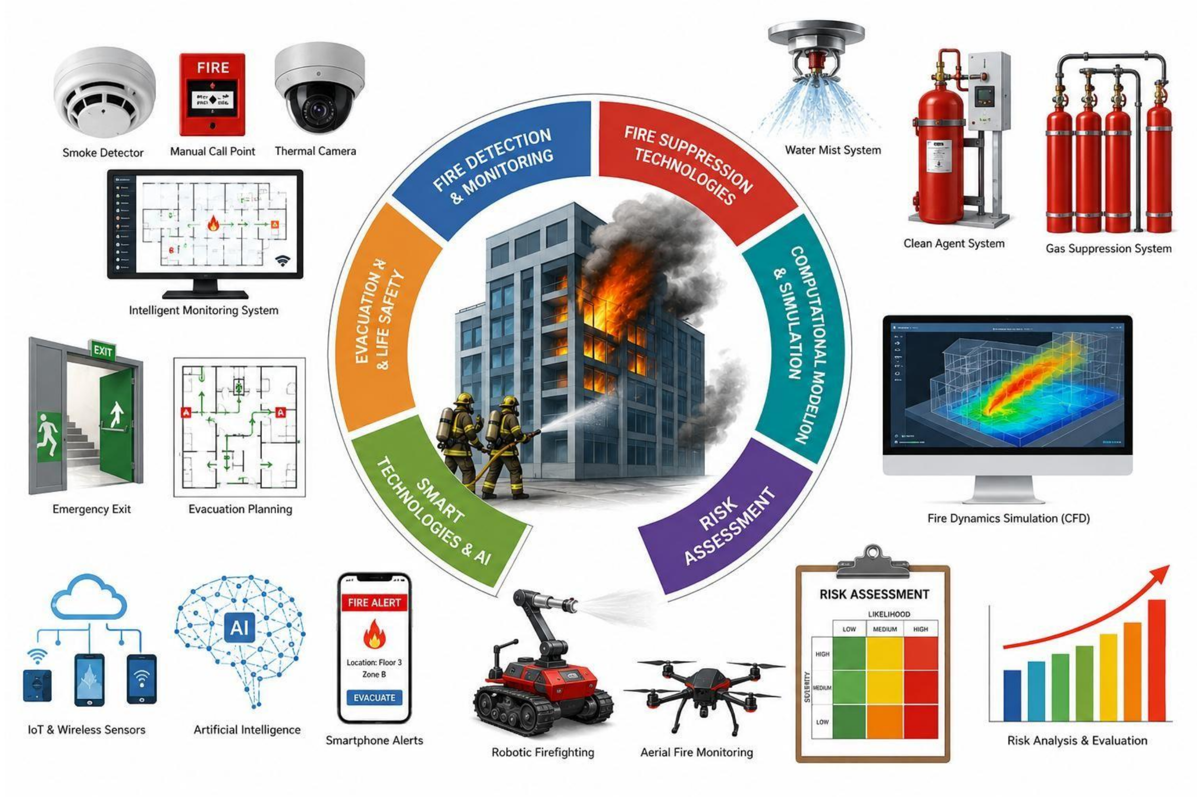


Figure 1: Fire safety systems and technologies overview

Overall, recent advances in fire dynamics and combustion analysis have greatly enhanced the understanding of fire behavior and improved the capability to design safer buildings, industrial facilities, and transportation systems. The integration of advanced computational modeling, experimental diagnostics, smart sensing technologies, and artificial intelligence continues to transform fire safety engineering into a more predictive, data-driven, and scientifically advanced discipline [14].

Advances in Fire Suppression Technologies

Fire suppression technologies have undergone remarkable advancements over the past few decades due to the increasing complexity of fire hazards in industrial, commercial, transportation, and residential environments. Modern fire suppression systems are designed not only to extinguish fires rapidly but also to minimize environmental impact, reduce water consumption, protect sensitive equipment, and improve operational reliability. The development of advanced suppression technologies has become increasingly important because traditional suppression methods are often insufficient for handling modern fire scenarios involving electrical systems, chemical storage facilities, lithium-ion batteries, hydrogen fuels, and data centers [6].

Conventional fire suppression systems mainly relied on water-based sprinkler systems, foam systems, carbon dioxide flooding, and dry chemical extinguishing agents. Although these systems remain widely used, they have certain limitations such as water damage, environmental concerns, limited effectiveness against specialized fires, and safety risks associated with gaseous agents. Therefore, researchers and fire protection engineers have focused on developing innovative suppression technologies that provide enhanced efficiency and sustainability [2].

One of the most significant advancements in modern fire suppression technology is the development of water mist systems. Water mist systems suppress fires through three primary mechanisms: cooling of flames and surfaces, displacement of oxygen through steam generation, and attenuation of radiant heat transfer [15]. Compared with conventional sprinkler systems, water mist technologies consume significantly less water, reduce collateral water damage, and provide faster suppression in enclosed spaces. These systems are increasingly used in marine vessels, aircraft hangars, tunnels, machinery spaces, power plants, heritage buildings, and data centers.

Another important advancement is the development of environmentally friendly clean agent suppression systems. Earlier halon-based suppression agents were highly effective but were phased out due to their ozone depletion potential and environmental hazards. Modern clean agents such as FK-5-1-12, inert gas blends, Novec-based agents, and fluoroketone compounds have emerged as safer alternatives [13]. These suppression systems extinguish fires either by reducing oxygen concentration or by absorbing heat from the combustion process. Clean agent systems are especially suitable for protecting sensitive electronic equipment, telecommunication systems, archives, museums, and server rooms because they leave no residue after discharge and minimize damage to critical assets.

Aerosol fire suppression technology has also gained considerable attention in recent years. Condensed aerosol systems generate fine solid particles and gaseous compounds that interrupt the chemical chain reactions responsible for combustion. Aerosol suppression units are compact, lightweight, cost-effective, and highly efficient in enclosed or confined spaces [1]. These systems are widely applied in electrical control panels, battery storage compartments, military vehicles, engine rooms, and industrial process equipment. Their ability to suppress fires without extensive piping infrastructure makes them attractive for retrofitting existing facilities.

Advances in foam-based suppression systems have further improved fire protection in flammable liquid storage facilities and petrochemical industries. Modern firefighting foams are designed to rapidly spread over fuel surfaces, suppress vapor release, and prevent re-ignition. Recent research has focused on developing fluorine-free foam formulations to address environmental and health concerns associated with per- and polyfluoroalkyl substances (PFAS) used in conventional aqueous film-forming foams (AFFF) [16]. Sustainable foam technologies are increasingly being adopted in airports, refineries, offshore platforms, and fuel storage terminals.

The emergence of lithium-ion battery technologies and renewable energy systems has introduced new challenges in fire suppression engineering. Battery fires are characterized by thermal runaway reactions that generate extremely high temperatures, flammable gases, and rapid fire propagation. Traditional suppression systems are often ineffective against such fires.

Consequently, advanced suppression methods involving water mist cooling, aerosol suppression, specialized chemical agents, and thermal management systems are being developed to mitigate battery fire hazards [10]. Research is also focused on suppression systems capable of preventing thermal runaway propagation between adjacent battery cells and modules.

Robotic firefighting technologies represent another major advancement in modern fire suppression engineering. Firefighting robots are designed to operate in hazardous environments where human intervention is dangerous or impossible. These robots are equipped with thermal imaging cameras, gas sensors, autonomous navigation systems, robotic arms, and high-capacity water or foam nozzles [5]. Modern robotic firefighting systems are increasingly used in chemical industries, nuclear facilities, underground mines, oil refineries, and military operations. Autonomous drones equipped with thermal sensors are also being developed for aerial firefighting, fire monitoring, and search-and-rescue operations.

Artificial intelligence and smart automation are now being integrated into suppression systems to improve response efficiency and reliability. Intelligent suppression systems can analyze realtime sensor data, predict fire growth patterns, activate suppression zones selectively, and optimize extinguishing agent discharge based on fire severity [9]. Smart building management systems integrated with AI-based fire suppression technologies provide faster response times and improved coordination between detection, suppression, ventilation, and evacuation systems.

Overall, recent advances in fire suppression technologies have significantly improved the effectiveness, sustainability, and adaptability of modern fire protection systems. The integration of smart sensing technologies, environmentally friendly suppression agents, robotic systems, and AI-driven automation has transformed traditional firefighting approaches into intelligent and highly efficient safety solutions. As industrial systems, energy infrastructures, and urban environments continue to evolve, further innovations in fire suppression technologies will remain essential for reducing fire risks and enhancing global fire safety standards [4].

Future Directions

The future of fire safety engineering is expected to be strongly influenced by rapid technological advancements, digital transformation, sustainability requirements, and the emergence of complex industrial and urban infrastructures. Modern fire safety strategies are gradually shifting from conventional reactive approaches toward predictive, intelligent, and fully integrated safety management systems. Future developments in fire safety engineering will focus on enhancing fire prevention, improving emergency response efficiency, minimizing environmental impacts, and increasing resilience against emerging fire hazards associated with smart cities, renewable energy systems, autonomous transportation, and advanced industrial processes [5].

One of the most promising future directions in fire safety engineering is the extensive integration of artificial intelligence (AI) and machine learning technologies. AI-based fire safety systems are expected to provide real-time predictive analytics by continuously processing data from smart sensors, surveillance systems, weather conditions, equipment performance records, and building management systems. Future intelligent fire safety platforms will be capable of identifying abnormal thermal patterns, predicting fire outbreaks before ignition occurs, and automatically initiating suppression and evacuation procedures [9]. Machine learning algorithms will further improve fire risk assessment models by learning from historical fire incidents and operational datasets, thereby increasing the accuracy of fire prediction and hazard identification.

The application of Internet of Things (IoT) technologies is also expected to expand significantly in future fire safety infrastructures. Smart buildings and industrial facilities will increasingly rely on interconnected wireless sensor networks capable of continuously monitoring temperature, smoke density, gas concentration, humidity, and structural conditions. These IoT-based systems will provide centralized monitoring and automated emergency communication through cloud-based platforms [13]. Real-time data transmission and remote accessibility will allow fire safety professionals to respond more rapidly and effectively during emergencies. Integration of IoT systems with smart city infrastructure will also improve coordination among emergency services, transportation systems, hospitals, and disaster management authorities. Future digital twin systems will enable dynamic fire risk assessment, predictive maintenance of fire protection systems, and advanced simulation of fire scenarios under changing environmental conditions [4]. Emergency responders will be able to visualize fire development, smoke movement, structural conditions, and occupant locations in real time, significantly improving decision-making during emergency operations.

Advanced robotics and autonomous firefighting technologies are also expected to play a major role in future fire suppression and rescue operations. Firefighting robots equipped with thermal imaging systems, artificial intelligence, autonomous navigation capabilities, and robotic manipulation systems will increasingly replace human firefighters in hazardous environments such as chemical plants, nuclear facilities, underground mines, offshore platforms, and military zones [16]. Autonomous aerial drones will be widely used for wildfire monitoring, fire mapping, smoke analysis, search-and-rescue missions, and delivery of emergency supplies. These robotic systems will reduce firefighter exposure to dangerous conditions while improving operational efficiency.

Future fire safety engineering will also emphasize sustainable and environmentally friendly suppression technologies. Conventional suppression agents associated with environmental pollution and greenhouse gas emissions will gradually be replaced by green suppression systems such as biodegradable foams, water mist technologies, and low-global-warmingpotential clean agents [6]. Research into nano-engineered fire suppression materials and smart extinguishing agents capable of adapting to specific fire conditions is expected to expand significantly. Sustainable fire protection solutions will become increasingly important due to stricter environmental regulations and global climate change concerns.

Another critical future research area involves fire safety challenges associated with renewable energy systems and advanced energy storage technologies. The increasing adoption of lithiumion batteries, hydrogen fuel systems, electric vehicles, and renewable power grids introduces unique fire hazards that require specialized engineering approaches. Future fire safety research will focus on improving thermal management systems, early gas detection technologies, battery fire suppression techniques, and hydrogen leak monitoring systems [10]. Computational modeling and experimental studies will continue to improve the understanding of thermal runaway mechanisms and explosion dynamics associated with these energy systems.

Human-centered fire safety engineering is also expected to evolve significantly in the future. Advanced evacuation models integrating behavioral psychology, crowd dynamics, and realtime sensor feedback will improve emergency planning and occupant safety. Smart evacuation systems using artificial intelligence, indoor positioning technologies, augmented reality guidance, and voice-assisted communication will provide adaptive evacuation instructions during emergencies [8]. Inclusive fire safety design for elderly individuals, children, and persons with disabilities will become an essential requirement in future building regulations and infrastructure planning.

The future of fire safety engineering will be characterized by intelligent automation, predictive analytics, sustainability, digital integration, and advanced materials innovation. The convergence of AI, IoT, robotics, computational modeling, and smart infrastructure technologies will significantly transform fire prevention, detection, suppression, and evacuation strategies. Continuous interdisciplinary research and international collaboration will be essential to address the emerging fire safety challenges associated with modern urbanization, industrialization, renewable energy systems, and global environmental changes [15].

Summary

Modern fire suppression technologies including water mist systems, clean agent suppression systems, aerosol-based extinguishing systems, and robotic firefighting technologies have enhanced firefighting efficiency while minimizing environmental impacts. Sustainable fire suppression agents and eco-friendly fire-resistant materials are increasingly being adopted to support green engineering practices.

The chapter highlighted the importance of structural fire engineering, evacuation modeling, and fire risk assessment in modern infrastructure design. Emerging technologies such as digital twins, autonomous firefighting robots, smart buildings, and AI-based predictive fire management systems are expected to transform future fire safety strategies.

Despite major technological advancements, challenges such as implementation costs, cybersecurity concerns, and fire risks associated with renewable energy systems continue to require further research and innovation. Overall, modern fire safety engineering plays a critical role in ensuring safer, smarter, and more resilient infrastructures worldwide.

References

1. Drysdale, D. (2011). *An introduction to fire dynamics* (3rd ed., pp. 1–512). John Wiley & Sons.
2. Hurley, M. J., *et al.* (2016). *SFPE handbook of fire protection engineering* (5th ed., pp. 1–3493). Springer.
3. Karlsson, B., & Quintiere, J. (2000). *Enclosure fire dynamics* (pp. 1–336). CRC Press.

4. McGrattan, K., *et al.* (2020). *Fire dynamics simulator user guide* (pp. 1–350). National Institute of Standards and Technology (NIST).
5. Fleming, J., & Moon, S. (2020). Application of artificial intelligence in fire detection systems. *Fire Technology, 56*(3), 1021–1042.
6. Wang, Q., Li, B., & Chen, X. (2022). Recent developments in water mist fire suppression systems. *Journal of Loss Prevention in the Process Industries, 75*(2), 104–120.
7. Buchanan, A. H., & Abu, A. (2017). *Structural design for fire safety* (2nd ed., pp. 1–450). John Wiley & Sons.
8. Fang, Z., Song, W., & Li, Z. (2018). Experimental study on human behavior in fire emergencies. *Safety Science, 110*(4), 240–249.
9. Sharma, R., *et al.* (2017). Artificial intelligence applications in industrial fire risk assessment. *Fire and Materials, 41*(6), 633–647.
10. Beyler, C. (2016). Fire hazard calculations for large open hydrocarbon fires. *Fire Safety Journal, 91*(2), 45–58.
11. Purser, D. (2017). Toxicity assessment of combustion products. In *SFPE handbook of fire protection engineering* (5th ed., pp. 2308–2428). Springer.
12. Abdelgawad, M., & Fayek, A. (2011). Risk management in the construction industry using combined fuzzy FMEA and fuzzy AHP. *Journal of Construction Engineering and Management, 137*(9), 1028–1036.
13. NFPA. (2023). *National Fire Protection Association codes and standards*. National Fire Protection Association.
14. Jiang, J., & Zhang, Y. (2021). Computational fluid dynamics modeling of smoke propagation in high-rise buildings. *Fire Safety Journal, 122*(5), 103–119.
15. Lataille, J. (2019). *Fire protection engineering in building design* (pp. 1–400). Elsevier.
16. United States Fire Administration. (2022). *Fire statistics and national fire data reports*.

MATHEMATICAL ANALYSIS OF MICROWAVE HEATING IN MULTIDIMENSIONAL GEOMETRIES WITH EXPONENTIAL SOURCE TERMS

P. Jeyabarathi and L. Rajendran
Department of Mathematics,
AMET, deemed to be University, Chennai, India
Corresponding author E-mail bsc.jeya@yahoo.com, raj_sms@rediffmail.com

Abstract

The present study develops an approximate analytical framework for modeling microwave heating in general geometries, including planar, cylindrical, and spherical configurations. The governing equation is formulated as a steady-state nonlinear reaction–diffusion equation with an internal exponential heat generation term arising from microwave energy absorption. An approximate analytical solution is obtained using the Akbari–Ganji method (AGM), which efficiently addresses the inherent nonlinearity of the system. The analytical results are validated against numerical simulations performed using MATLAB over a wide range of governing parameters, such as the dimensionless microwave absorption coefficient and thermal diffusivity. The comparison shows excellent agreement, confirming the accuracy and robustness of the proposed approach. Furthermore, parametric analysis reveals that increasing internal heat generation significantly intensifies temperature gradients within the medium. The developed model provides useful insight into microwave-assisted heating processes and serves as an efficient tool for analyzing nonlinear thermal systems in engineering applications.

Keywords: Nonlinear Equation, Reaction Diffusion, Microwave Heating, General Geometries, Akbari–Ganji Method, Numerical Simulation.

1. Introduction

Microwaves are a form of electromagnetic radiation characterized by oscillating electric and magnetic fields that propagate through space. Within the electromagnetic spectrum, they occupy a distinct region of the radio frequency band, lying between infrared radiation and conventional radio waves. Their ability to penetrate materials and generate volumetric heating makes them particularly important in thermal processing and energy transfer applications.

Microwave heating in solid media has attracted considerable attention due to its relevance in industrial and engineering processes. Heat conduction in such systems, particularly in general geometries including slab, cylindrical, and spherical configurations, can be modeled using nonlinear reaction–diffusion equations with internal heat generation. Classical heat transfer analyses [1,2] and diffusion studies [3] provide the theoretical foundation for formulating these models. In microwave heating, the internal heat source arises from electromagnetic energy absorption, which typically decays exponentially within the material. This leads to a nonlinear governing equation in which temperature-dependent absorptivity introduces significant analytical complexity.

To facilitate analysis, the governing equation is transformed into a dimensionless nonlinear boundary value problem, allowing systematic investigation of key parameters such as thermal absorptivity and electric field attenuation. However, due to the strong nonlinearity and spatially varying source term, exact solutions are generally not attainable. Consequently, approximate analytical methods become essential for obtaining tractable solutions.

Among the available techniques, the Akbari–Ganji Method (AGM) [4] provides an efficient framework for constructing accurate analytical approximations to nonlinear differential equations. This method yields closed-form expressions that closely agree with numerical results, thereby offering valuable insight into the thermal behavior of microwave-heated systems. Additionally, the Hermite-approximation Padé method [5–6] has emerged as a powerful semi-analytical tool and benchmarking approach for nonlinear problems. It has been successfully applied in various scientific and engineering contexts and serves as a reliable means of validating numerical solutions. For instance, Makinde [7] employed this method to analyze steady-state nonlinear temperature fields.

In the present study, the nonlinear reaction–diffusion model governing microwave heating is formulated in a unified framework applicable to general geometries. An approximate analytical solution for the temperature distribution is derived using the Akbari–Ganji Method, enabling accurate prediction of thermal behavior under varying geometric and physical parameters.

Nomenclature

Symbol	**Meaning**
$u(x) = \frac{T}{T_0}$	Dimensionless temperature
$x = \frac{y}{a}$	Dimensionless spatial coordinate
$\lambda = \frac{Ea^2 T_0^{n-1}}{\alpha}$	Thermal absorptivity parameter
$k = \alpha\beta$	Electric field decay parameter
$m \in \{0,1,2\}$	Geometry parameter (slab, cylindrical, spherical)

2. Mathematical Formulation of the Problem

The steady-state nonlinear reaction–diffusion equation incorporating a source term to represent thermal effects is given by [7]

$$\alpha\, T''(y) + \frac{m}{x} T'(y) + E\, e^{-\beta y} T(y)^n = 0 \qquad (1)$$

The boundary conditions are

$$T'(0) = 0, T(a) = T_0 \qquad (2)$$

$T(y)$denotes the temperature distribution, y represents the spatial coordinate measured in the normal direction, E is the amplitude of the incident radiation, α is the thermal diffusivity constant, β is the electric field decay rate, n is the thermal absorptivity index and m is the general geometries. Furthermore, $a$denotes the half-width of the slab, and T_0 is the prescribed wall temperature.

By introducing Eq. (1) in the following dimensionless form

$$x = \frac{y}{a}, u = \frac{T}{T_0}, \lambda = \frac{E\, a^2 T_0^{n-1}}{\alpha}, k = \alpha\, \beta \quad (3)$$

Governing Dimensionless Model

$$\frac{1}{x^m}\frac{d}{dx}\left(x^m \frac{du}{dx}\right) + \lambda e^{-kx} u^n = 0 \quad (4)$$

Boundary Conditions

$$\frac{du}{dx}\Big|_{x=0} = 0, u(1) = 1 \quad (5)$$

$u(x)$ denotes the dimensionless temperature field, λ represents the thermal absorptivity parameter, k is the electric field decay parameter, and x is the dimensionless spatial coordinate.

3. Approximate Analytical Expression of the Temperature Field Using Akbari–Ganji Method

According to the Akbari–Ganji Method (AGM), the approximate solution is assumed in the form of a trial function containing unknown constant coefficients. This trial function is constructed such that it satisfies both the governing differential equation and the prescribed boundary conditions. Accordingly, for Eq. (4), a quadratic polynomial trial solution is considered as

$$u(x) = \sum_{i=0}^{2} u_i\, x^i = u_0 + u_1 x + u_2 x^2 \quad (6)$$

where u_0, u_1 and u_2 are constants. Using the boundary conditions (5), we get

$$u_1 = 0, \; u_0 = -u_2 \quad (7)$$

Now define the function F by

$$F(x = 0.5) = u''(x) + \frac{m}{x} u'(x) + \lambda\, e^{-kx} u(x)^n = 0 \quad (8)$$

$$2\, u_2(1 + m) + \lambda\, e^{-0.5\, k}(1 + 0.75\, u_2)^n = 0 \quad (9)$$

Using eq. (10), the eq. (7) can be rewritten as follows:

$$u(x) = 1 + u_2(x^2 - 1) \quad (10)$$

The unknown parameter u_2 can be obtained by solving the equation (11) using Matlab

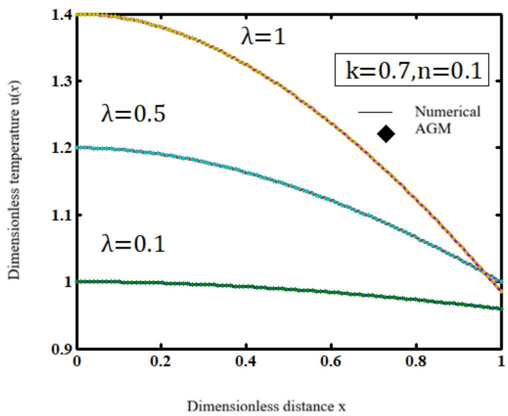


Figure 1: Comparison of analytical expression of temperature field(m=0) $u(x)$ with numerical results for various value of parameter λ using eq. (11)

Figure 1 represents the effects of the thermal absorptivity parameter λon the temperature profile for various reaction mechanisms. From the figure, it is observed that the temperature increases with increasing thermal absorptivity. This behavior indicates that higher values of λ enhance the absorption of incident radiation, leading to greater internal heat generation within the medium. Consequently, the dimensionless temperature $u(x)$attains higher values throughout the domain. The effect is more pronounced near the surface region, while all profiles satisfy the boundary condition at $x = 1$. Furthermore, the excellent agreement between the AGM analytical solution and the numerical results validates the accuracy of the proposed approach.

Conclusion

In this study, microwave heating has been analytically investigated by incorporating thermal absorptivity effects in generalized geometries, including slab, cylindrical, and spherical configurations. The governing steady-state nonlinear reaction–diffusion equation describing the temperature distribution was formulated in a unified dimensionless form. An approximate analytical solution was obtained using the Akbari–Ganji Method (AGM), enabling accurate prediction of temperature profiles under varying geometric conditions.

References

1. Carslaw, H. S., & Jaeger, J. C. (1959). *Conduction of heat in solids*. Oxford University Press.
2. Metaxas, A. C., & Meredith, R. J. (1983). *Industrial microwave heating*. Peter Peregrinus Ltd.
3. Crank, J. (1975). *The mathematics of diffusion*. Oxford University Press.
4. Akbari, M., & Ganji, D. D. (2014). Application of the Akbari–Ganji method to nonlinear differential equations. *International Journal of Engineering and Applied Sciences*.
5. Baker, G. A., Jr., & Graves-Morris, P. (1996). *Padé approximants* (2nd ed.). Cambridge University Press.
6. Makinde, O. D. (1999). Extending the utility of perturbation series in problems of laminar flow in a porous pipe and a diverging channel. *Journal of the Australian Mathematical Society, Series B, 41*, 118–128.
7. Tourigny, Y., & Drazin, P. G. (2000). The asymptotic behaviour of algebraic approximants. *Proceedings of the Royal Society of London A, 456*, 1117–1137.
8. Makinde, O. D. (2007). Solving microwave heating model in a slab using Hermite–Padé approximation technique. *Applied Thermal Engineering, 27*, 599–603.

MICROPLASTIC CONTAMINATION IN FRESHWATER ECOSYSTEMS: DETECTION METHODS, ECOLOGICAL IMPACT, AND MITIGATION STRATEGIES

Rajesh Kumar Mishra[1], Divyansh Mishra[2] and Rekha Agarwal[3]

[1]ICFRE-Tropical Forest Research Institute

(Ministry of Environment, Forests & Climate Change, Govt. of India)

P.O. RFRC, Mandla Road, Jabalpur, MP-482021, India

[2]Department of Artificial Intelligence and Data Science

Jabalpur Engineering College, Jabalpur (MP)

[3]Government Science College, Jabalpur, MP, India- 482 001

Corresponding author E-mail: rajeshkmishra20@gmail.com, rajesh.mishra0701@gov.in, divyanshspps@gmail.com, rekhasciencecollege@gmail.com

Abstract

Microplastic contamination has emerged as a critical environmental challenge of the twenty-first century, with freshwater ecosystems increasingly recognized as both significant pathways and ultimate repositories for plastic debris. This chapter provides a comprehensive synthesis of the current state of knowledge on microplastic pollution in rivers, lakes, and wetlands. We examine the origins and classification of micro plastics — distinguishing primary from secondary sources — and trace the diverse pathways through which these particles enter freshwater systems. A detailed review of detection and analytical methodologies is presented, covering traditional microscopy and spectroscopy techniques alongside emerging approaches such as AI-assisted imaging, biosensors, and IoT-enabled monitoring platforms. The ecological consequences of microplastic exposure are critically assessed at multiple biological levels, from cellular oxidative stress responses in individual organisms to population-level reproductive impairment and trophic transfer through freshwater food webs. Finally, mitigation strategies are evaluated across three dimensions: engineered removal technologies in wastewater treatment (including membrane bioreactors, electrocoagulation, and photocatalytic degradation), nature-based and bioremediation solutions, and governance frameworks including extended producer responsibility and circular economy policies. The chapter concludes by identifying critical knowledge gaps and proposing a research agenda to address them. Understanding and managing microplastic contamination in freshwater ecosystems is essential for safeguarding biodiversity, human drinking water supplies, and long-term ecosystem services.

Keywords: Micro Plastics, Freshwater Pollution, Spectroscopy, Ecotoxicology, Wastewater Treatment, Circular Economy, Biodiversity.

1. Introduction

The pervasive proliferation of plastic materials since the mid-twentieth century has generated one of the most intractable environmental problems of our era. Global plastic production has surpassed 380 million tonnes annually, and a significant proportion of this output enters the natural environment through inadequate waste management, atmospheric deposition, and deliberate or accidental discharge (Debnath *et al.,* 2023). Within the broader spectrum of plastic pollution, micro plastics — particles typically defined as plastic fragments smaller than 5 millimetres — have attracted intense scientific scrutiny owing to their ubiquity, persistence, and capacity to interact with biological systems at multiple scales (Thompson *et al.,* 2004; Rochman *et al.,* 2019).

Although initial research attention was directed largely toward marine environments, growing evidence has established that freshwater ecosystems — rivers, lakes, reservoirs, wetlands, and groundwater — function simultaneously as conduits transporting plastic debris to the oceans and as significant long-term sinks in their own right (Horton *et al.,* 2017). The ecological stakes are considerable: freshwater systems support an estimated 10% of all known species, provide drinking water for billions of people, sustain fisheries and agriculture, and underpin the water cycle that regulates climate. Contamination of these systems by micro plastics thus carries implications that extend far beyond the aquatic organisms directly exposed.

The United Nations has formally recognized micro plastics as emerging contaminants of global concern, and the World Economic Forum has projected that, without decisive intervention, plastic contamination levels in aquatic environments will double by 2030 (WEF, 2022). These projections lend urgency to the scientific enterprise of understanding the sources, behaviour, ecological effects, and remediation of micro plastics in freshwater systems.

This chapter provides a structured, evidence-based synthesis of that enterprise. Section 2 defines micro plastics and examines their sources and environmental pathways. Section 3 reviews detection and characterization methodologies in depth. Section 4 analyses ecological impacts across biological organization levels. Section 5 evaluates mitigation and remediation strategies. Section 6 surveys the regulatory landscape. Section 7 identifies knowledge gaps and research priorities. The chapter draws on peer-reviewed literature spanning over two decades, with particular emphasis on studies published between 2018 and 2025.

2. Sources, Classification, and Environmental Pathways

2.1 Classification of Micro plastics

Micro plastics are conventionally classified into two broad categories based on their origin. Primary micro plastics are manufactured at small sizes for specific industrial or consumer applications: these include micro beads incorporated into personal care products and cosmetics, plastic pellets (nurdles) used as raw material in polymer manufacturing, and synthetic fibres shed from textiles during washing. Secondary micro plastics, by contrast, arise from the fragmentation of larger plastic items under the cumulative influence of ultraviolet radiation, mechanical abrasion, wave action, thermal cycling, and biological processes (Andrady, 2017; Zhao *et al.,* 2015).

Further classification by morphology distinguishes fragments, fibres, films, foams, pellets, and beads — each with distinct transport dynamics, bioavailability, and biological effects. Polymer composition is equally relevant: common freshwater micro plastics include polyethylene (PE), polypropylene (PP), polystyrene (PS), polyethylene terephthalate (PET), polyamide (nylon), and polyvinyl chloride (PVC), each carrying different sorption capacities for environmental contaminants (Avio *et al.,* 2017).

Table 1: Classification of Micro plastics by Origin, Morphology, and Common Polymer Types

Category	Sub-type	Primary Sources	Common Polymers
Primary	Microbeads	Cosmetics, personal care products	PE, PP, PS
Primary	Industrial pellets (nurdles)	Polymer manufacturing, transport	PE, PP, PET
Primary	Synthetic fibres	Textile washing, apparel	Polyamide, PET, acrylic
Secondary	Fragments	Degradation of bottles, packaging	PE, PP, PS, PVC
Secondary	Films	Degradation of agricultural films, bags	PE, PP
Secondary	Foams	Degradation of packaging foam	Expanded PS

2.2 Pathways into Freshwater Ecosystems

Micro plastics reach freshwater bodies through a diverse array of pathways. Wastewater treatment plant (WWTP) effluents constitute one of the most quantitatively significant routes: even after treatment, effluents may discharge millions of micro plastic particles per day into receiving water bodies, because conventional treatment infrastructure was not designed to target particles at this scale (Prata, 2018; Ahmed *et al.,* 2024). Storm water runoff collects micro plastics deposited on impervious urban surfaces — road markings, tyre wear particles, litter — and transports them directly to streams and rivers (Baun *et al.,* 2020). Atmospheric deposition of airborne micro plastics, including synthetic fibres from laundering, has been documented even in remote alpine lakes, demonstrating that no freshwater body is truly isolated from anthropogenic plastic contamination (Allen *et al.,* 2019). Agricultural practices represent an under-appreciated pathway: the application of sewage sludge as fertilizer introduces micro plastics-rich material to soils, from which particles migrate to drainage systems and rivers. Plastic mulch films and irrigation from polluted surface water further contribute to agricultural loading (Nizzetto *et al.,* 2016). Industrial discharges, accidental spills during pellet transport, and the direct littering of riverbanks and lakeshores complete the principal exposure routes.

Once in the freshwater environment, micro plastics undergo complex redistribution governed by particle density, shape, and size; hydrological regime; and biogeochemical interactions. Buoyant particles (PE, PP) tend to be retained at the water surface or transported laterally, while denser plastics (PET, PVC, PS) settle into sediments, where microplastic concentrations can reach values comparable to those found in the world's most contaminated marine sediments (Horton & Dixon, 2018). Floodplain inundation events can remobilise sediment-bound plastics over large areas, effectively re-activating previously deposited contaminant burdens.

3. Detection and Analytical Methods

The reliable detection, quantification, and characterization of micro plastics in freshwater matrices present substantial methodological challenges arising from the heterogeneity of particle sizes, morphologies, and polymer types; the complexity of environmental sample matrices; and the absence of universally adopted standardized protocols (Carvalho *et al.,* 2021). A tiered analytical workflow — sampling, pre-treatment, visual identification, and chemical characterization — is generally recommended to ensure reproducibility and comparability across studies.

3.1 Sampling Strategies

Sampling design must account for the three-dimensional distribution of micro plastics across water surface, subsurface water column, and sediment compartments. Manta trawls equipped with fine-mesh nets (typically 333 µm) are widely used for surface water collection in rivers and lakes, while bulk water sampling with subsequent filtration is preferred when smaller size fractions (< 333 µm) are targeted. Sediment sampling employs grab samplers, corers, or stainless-steel spades. Biotic matrices — fish gut contents, invertebrate tissue — require tissue homogenisation followed by chemical digestion to isolate embedded particles. For organisms specifically, enzymatic digestion using proteinase K or hydrogen peroxide is preferred to preserve polymer integrity while eliminating biological material (Carvalho *et al.,* 2021).

3.2 Visual Identification and Microscopy

Visual identification under stereo or optical microscopy remains the most common initial screening step, providing particle counts, morphological classification, and colour distribution. However, visual methods are prone to observer bias and cannot definitively confirm polymer identity. Fluorescent dyes — most notably Nile Red — have been introduced as a cost-effective screening tool that tags hydrophobic plastic surfaces, enabling rapid fluorescence microscopy-based identification even at particle sizes below 100 µm (Maes *et al.,* 2017). Scanning Electron Microscopy (SEM) coupled with Energy Dispersive X-ray Spectroscopy (SEM-EDX) offers high-resolution surface morphology and elemental composition data, but remains time-intensive and unsuitable for large-volume environmental monitoring.

3.3 Spectroscopic Methods

Chemical characterization is essential for unambiguous polymer identification. Fourier-Transform Infrared Spectroscopy (FTIR) in attenuated total reflectance (ATR) mode generates molecular fingerprint spectra matched against reference libraries, enabling reliable identification of polymer types down to particle sizes of approximately 20 µm. Micro-FTIR imaging extends this capability to individual particles within complex mixtures. Raman spectroscopy offers superior spatial resolution (theoretical limit ~1 µm) and is applicable to wet samples without prior drying, making it particularly suited to aquatic micro plastic analysis; its principal limitation is fluorescence interference from organic contaminants and coloured particles, which can be mitigated by surface-enhanced Raman scattering (SERS) techniques (Lohmann *et al.,* 2021).

Pyrolysis-Gas Chromatography-Mass Spectrometry (Py-GC-MS) provides bulk chemical characterization by thermally decomposing plastic polymers and analysing the resulting volatile compounds, enabling mass-based quantification of polymer classes in complex matrices. While destructive and incompatible with morphological information, Py-GC-MS is particularly valuable for detecting small nanoplastics and for quantifying low concentrations in environmental samples (Fischer & Scholz-Böttcher, 2017).

3.4 Emerging Detection Technologies

The limitations of conventional methods — cost, throughput, and skill requirements — have stimulated rapid innovation in detection technology. AI-powered imaging systems employing convolutional neural networks (CNNs) can automatically classify microplastic morphologies and polymer types from hyperspectral or RGB images with accuracy exceeding 90% in controlled laboratory settings, dramatically accelerating processing times compared to manual analysis (Thangagiri & Sivakumar, 2024; NIH/PMC, 2024). These systems are increasingly being coupled with automated sample processing pipelines to enable true high-throughput monitoring.

Biosensor technologies represent another frontier with considerable promise. Electrochemical biosensors exploit the interaction between plastic particles and functionalised electrode surfaces to generate quantifiable electrical signals; optical biosensors using localised surface plasmon resonance (LSPR) or fluorescence quenching have demonstrated detection sensitivities in the nanogram-per-litre range for specific polymer types (Praveena *et al.,* 2024). IoT-enabled sensor networks, integrating low-cost optical sensors, microcontrollers, and cloud-based data analytics, are beginning to be deployed for near-real-time microplastic monitoring in freshwater systems, offering the spatial and temporal resolution needed to understand dynamic transport processes (Sharma *et al.,* 2024).

Table 2: Comparison of Principal Microplastic Detection Methods

Method	Size Limit	Polymer ID?	Throughput	Key Limitations
Visual microscopy + Nile Red	~20 µm	No	Medium	False positives; no chemical ID
FTIR / µ-FTIR	~20 µm	Yes	Low–Medium	Requires dry sample; time-intensive
Raman spectroscopy	~1 µm	Yes	Low	Fluorescence interference
Py-GC-MS	Bulk/nanoscale	Yes (bulk)	Low	Destructive; no particle morphology
SEM-EDX	~100 nm	Partial	Very Low	Expensive; vacuum conditions
AI-assisted hyperspectral imaging	~5 µm	Yes	High	Requires large training datasets
Biosensors (electrochemical/optical)	~100 nm	Partial	High	Selectivity; matrix effects

4. Ecological Impacts

The ecological consequences of microplastic contamination in freshwater ecosystems are multifaceted and operate across levels of biological organisation, from sub-cellular biochemical disruption to ecosystem-level shifts in energy flow and trophic dynamics. While much foundational toxicological work has been conducted on marine organisms, a growing body of freshwater-specific research is revealing that freshwater species are not inherently less sensitive to microplastic exposure (Horton & Dixon, 2018; Berlino *et al.,* 2023).

4.1 Physical and Physiological Effects on Organisms

Physical ingestion of micro plastics has been documented across virtually all taxonomic groups inhabiting freshwater: fish, amphibians, invertebrates, waterfowl, and riparian mammals. Ingestion may be active (mistaking plastic particles for food items based on size, colour, or odour) or passive (incidental uptake with water or prey). Once ingested, micro plastics can physically obstruct the gastrointestinal tract, reducing digestive efficiency and inducing a false sense of satiation that suppresses normal feeding behaviour, thereby reducing energy intake and compromising growth and reproductive output (Gallitelli *et al.,* 2021; Berlino *et al.,* 2023).

In fish, histopathological studies have demonstrated structural damage to intestinal epithelium, hepatic tissue, gill lamellae, and neural tissue following microplastic exposure, with severity depending on particle size, shape, polymer type, concentration, and exposure duration (Jabeen *et al.,* 2017). Microplastic-induced oxidative stress — evidenced by elevated reactive oxygen species (ROS) production, lipid peroxidation, and upregulation of antioxidant enzymes — has been consistently reported and is considered a primary mechanism of sub-lethal toxicity (Chen *et al.,* 2020). Effects on the endocrine system, including disruption of thyroid hormone regulation and reproductive steroid synthesis, have been linked to both the plastic polymers themselves and to sorbed chemical additives (plasticisers, flame retardants, UV stabilisers) that leach from particle surfaces (Rochman *et al.,* 2019).

In aquatic invertebrates — including Daphnia (water fleas), chironomid midges, gammarid amphipods, and caddisfly larvae — microplastic exposure causes reductions in feeding rate, growth, reproduction, and survival. A 2025 meta-analysis of 47 studies examining combined microplastic and temperature stress effects on freshwater invertebrates concluded that elevated temperature significantly exacerbates microplastic toxicity, with reproductive endpoints showing particularly strong negative synergistic responses — a finding with stark implications in the context of climate change (Environmental Sciences Europe, 2025). For filter-feeding bivalves and passive suspension feeders, the challenge of discriminating plastic particles from natural suspended organic matter is acute; research on freshwater bivalves indicates that chronic sub-lethal exposure alters immune function and increases susceptibility to pathogens.

4.2 Micro plastics as Vectors for Co-contaminants

A critically important ecological dimension of micro plastics is their function as vectors for co-contaminants. Due to their hydrophobic surface chemistry and large surface-area-to-volume ratio, microplastic particles sorb persistent organic pollutants (POPs) including polychlorinated biphenyls (PCBs), polycyclic aromatic hydrocarbons (PAHs), dichlorodiphenyltrichloroethane (DDT) and its metabolites, and organochlorine pesticides, often at concentrations several orders of magnitude above those in surrounding water (Avio *et al.,* 2017). Toxic metals including cadmium, lead, and chromium also accumulate on microplastic surfaces through surface complexation. When ingested by organisms, the combination of desorption kinetics in gut conditions, the physical damage to gut lining that may enhance uptake, and the leaching of plastic additives creates a complex chemical exposure scenario that is difficult to separate from the physical effects of the particles themselves.

The 'plastisphere' — the distinct microbial community that colonises microplastic surfaces in aquatic environments — represents an additional ecological concern. Plastisphere biofilms have been found to harbour potentially pathogenic bacteria and antibiotic resistance genes at significantly higher densities than surrounding water, raising the prospect that micro plastics act as 'Trojan horses' facilitating pathogen dispersal and the spread of antimicrobial resistance through freshwater networks (Arias-Andres *et al.,* 2018).

4.3 Food Web and Ecosystem-Level Effects

Trophic transfer of micro plastics — and of associated contaminants — through freshwater food webs is now well-established. Zooplankton that ingest plastic-contaminated phytoplankton are themselves consumed by invertebrates and larval fish, which in turn are predated by larger fish, waterfowl, and riparian mammals including otters and mink. Biomagnification of sorbed POPs along these food chains can result in tissue concentrations at upper trophic levels that substantially exceed those in primary consumers (Rochman *et al.,* 2019). The density of microplastic particles in some polluted river systems has been reported to exceed that of living zooplankton, fundamentally altering the food available to planktivorous organisms (Horton & Dixon, 2018).

At the ecosystem level, microplastic contamination has the potential to disrupt key processes including primary production (by reducing light penetration and affecting phytoplankton photosynthesis), decomposition (by interfering with shredder invertebrates that process leaf litter and organic matter), and nutrient cycling (through effects on benthic invertebrate communities and microbial biogeochemical pathways). A meta-analysis of functional trait responses concluded that metabolism, growth, and reproduction of benthic organisms were negatively affected, and that responses differed by trophic level, suggesting negative effects on energy transfer through food webs (Berlino *et al.,* 2023).

Bayesian species sensitivity distribution (SSD) modelling has estimated median hazardous concentrations (HC5) for freshwater microplastic spheres or fragments in the range of 0.02–2 μg/L, while fibres — which are particularly abundant in freshwater systems receiving textile wastewater — show HC5 values roughly an order of magnitude lower, indicating substantially greater ecotoxicological concern (Mehinto *et al.,* 2022). These thresholds have been exceeded in numerous documented freshwater systems, indicating that ecologically relevant exposures are already occurring at current contamination levels.

5. Mitigation and Remediation Strategies

Addressing microplastic contamination in freshwater ecosystems requires a multi-pronged strategy encompassing upstream prevention (reducing plastic production and emissions), in-system treatment (removing micro plastics from wastewater and water supplies), and environmental remediation (recovering particles already distributed in the environment). No single technological solution is sufficient; integrated approaches combining engineering, biology, and governance are necessary.

5.1 Wastewater Treatment Technologies

Conventional wastewater treatment plants (WWTPs) — designed primarily to remove biological oxygen demand and nutrients — achieve variable and often insufficient microplastic removal, particularly for particles below 10 μm and fibres. Despite this, primary and secondary treatment collectively can remove 70–95% of micro plastics present in influent, predominantly by sedimentation and biological floc formation (Ahmed *et al.,* 2024). The principal limitation is that the remaining fraction — typically small particles and fibres that remain in suspension — is discharged in effluent at concentrations that, multiplied by the enormous daily volumes processed by large WWTPs, represent millions to billions of particles released daily to receiving water bodies.

5.1.1 Membrane Filtration

Membrane-based technologies offer the most reliable barrier for fine microplastic removal. Ultrafiltration (UF) membranes with pore sizes of 0.01–0.1 μm can theoretically retain all particles above this threshold. Membrane Bioreactors (MBRs), combining biological treatment with membrane separation, consistently achieve microplastic removal efficiencies exceeding 99% in pilot and full-scale studies, while also reducing pharmaceutical micropollutants (Lares *et al.,* 2018). The primary barriers to universal MBR adoption are capital cost, energy consumption (3–10 times that of conventional activated sludge), and membrane fouling requiring regular cleaning with chemical agents. Nanofiltration and reverse osmosis provide even finer filtration but are currently reserved for drinking water treatment due to cost.

5.1.2 Rapid Sand Filtration and Coagulation-Flocculation

Rapid sand filtration (RSF) is a widely employed and relatively cost-effective tertiary treatment step that has demonstrated microplastic removal efficiencies of 74–97% for particles above 10 μm, with biochar-integrated sand filters achieving over 95% removal through enhanced surface adsorption mechanisms (Hou *et al.,* 2021; Wang *et al.,* 2020). Coagulation-flocculation using aluminium sulphate or polyaluminium chloride destabilises colloidal micro plastics, enabling their removal by sedimentation or flotation. Studies report 70–90% removal efficiency, with the added benefit of removing co-associated contaminants sorbed to plastic surfaces. However, the process generates substantial sludge volumes that may themselves contain concentrated micro plastics alongside heavy metals and organic pollutants, creating a secondary waste management challenge (Ahmed *et al.,* 2024).

5.1.3 Electrocoagulation and Advanced Oxidation Processes

Electrocoagulation employs sacrificial metal electrodes to generate coagulant cations in situ, avoiding chemical dosing while achieving comparable or superior removal efficiencies. Studies using aluminium-aluminium electrode configurations have reported polystyrene removal exceeding 90% under optimised conditions (Subair *et al.,* 2023). Advanced oxidation processes

(AOPs) including photocatalytic degradation (using TiO_2 under UV irradiation), ozonation, and the electro-Fenton process generate highly reactive hydroxyl radicals capable of mineralising plastic polymers. The electro-Fenton process employing TiO_2/graphite electrodes has achieved up to 75% dechlorination and 56% weight loss of PVC, representing a pathway to actual polymer destruction rather than particle removal and concentration in a secondary waste stream (Bule Možar *et al.,* 2024). The high reactivity of photocatalysts allows for complete polymer mineralisation under optimised conditions, offering particular promise for treating recalcitrant micro plastics in industrial effluents.

5.2 Biological and Nature-Based Remediation

Bioremediation approaches harness the metabolic capabilities of microorganisms and algae to remove or degrade micro plastics. Microalgae, including Chlorella vulgaris and Spirulina platensis, have demonstrated the capacity to aggregate micro plastics through biosorption and extracellular polymeric substance (EPS) secretion, forming heteroaggregates that settle more rapidly from suspension (Gao *et al.,* 2023). The formation of these plastisphere-microalgae aggregates, while environmentally occurring, can be deliberately engineered in photobioreactors to serve as a low-energy, low-cost tertiary treatment stage. Constructed wetlands incorporating emergent macrophytes, microbial biofilms, and filtration media have been evaluated as passive polishing stages for WWTP effluent, with removal efficiencies of 60–80% for micro plastics reported in pilot studies, though fibres remain challenging to capture.

Plastic-degrading microorganisms, including species of Ideonella sakaiensis, Pseudomonas, and Bacillus, have been identified as capable of enzymatically depolymerising PET and certain polyolefins, though the timescales for complete mineralisation under environmental conditions remain orders of magnitude longer than practical remediation requirements (Yoshida *et al.,* 2016). Accelerating enzymatic degradation through bioaugmentation, metabolic engineering, and the use of engineered plastic-degrading enzymes (PETases, MHETases) represents an active research frontier.

5.3 Source Reduction and Product Design

Upstream prevention — reducing the quantity of plastic entering the environment — is ultimately more cost-effective and ecologically sound than downstream remediation. Several product-level interventions have demonstrated measurable impacts. Washing machine filters designed to capture synthetic fibres before they enter wastewater have shown promise in pilot trials, with capture efficiencies of 60–90% for fibres above 50 µm. The development of biodegradable or bio-based polymer alternatives for single-use applications can reduce the long-term persistence of particles that do escape into the environment. Reformulation of personal care products to eliminate intentionally added microbeads — already legislated in several jurisdictions including the United States, United Kingdom, and European Union — has removed one quantitatively significant primary microplastic source from wastewater streams.

Table 3: Summary of Microplastic Mitigation Technologies — Efficacy, Cost, and Applicability

Technology	Removal Efficiency	Relative Cost	Scalability	Primary Limitation
Membrane Bioreactor (MBR)	>99%	High	Medium	Energy use; membrane fouling
Rapid sand filtration	74–97%	Low–Medium	High	Limited for very small particles
Coagulation-flocculation	70–90%	Medium	High	Sludge disposal challenges
Electrocoagulation	>90% (PS)	Medium	Medium	Energy cost; electrode maintenance
Photocatalytic degradation (AOP)	Variable	High	Low–Medium	Energy; reactor design complexity
Constructed wetlands	60–80%	Low	Medium	Land footprint; seasonal variation
Microalgae bioreactors	50–80%	Medium	Low–Medium	Optimisation; harvesting efficiency
Washing machine fibre filters	60–90% fibres	Low	High (household)	Effectiveness for nano-fibres

6. Regulatory Frameworks and Policy Dimensions

The governance of microplastic pollution exists at the intersection of multiple regulatory domains — chemicals management, waste regulation, water quality standards, and product design requirements — and has evolved unevenly across jurisdictions. In the European Union, the Registration, Evaluation, Authorisation and Restriction of Chemicals (REACH) regulation has been used to restrict the intentional use of micro plastics in consumer products; the EU's Single-Use Plastics Directive (2019/904) targets the most problematic single-use plastic items that become secondary microplastic precursors. The European Green Deal and the EU Action Plan for a Circular Economy provide a broader framework for redesigning production and consumption systems to reduce plastic waste at source.

At the international level, negotiations toward a legally binding Global Plastics Treaty under the auspices of the United Nations Environment Programme (UNEP) have proceeded through multiple Intergovernmental Negotiating Committee (INC) sessions. The failure to reach a final agreement at INC-5 in Busan, South Korea in December 2024 highlighted persistent tensions between nations favouring mandatory production caps and those preferring voluntary measures, and underscored the challenges of achieving the legally binding global framework that scientists argue is necessary given the transboundary nature of microplastic pollution (Environmental Sciences Europe, 2025).

Extended Producer Responsibility (EPR) schemes, which assign the cost of end-of-life plastic management to producers rather than taxpayers or the environment, are increasingly incorporated into national plastic strategies. When well-designed, EPR provides economic incentives for producers to redesign products for recyclability, reduce packaging weight, and invest in collection infrastructure. Ecolabelling schemes for low-microplastic textiles and biodegradable alternatives can guide consumer choices, while industry standards for plastic pellet containment (such as the Operation Clean Sweep initiative) address primary microplastic losses at manufacturing and distribution points.

7. Research Gaps and Future Directions

Despite rapid scientific progress, several critical knowledge gaps constrain our ability to fully assess and manage microplastic contamination in freshwater ecosystems. The following priorities are identified based on a synthesis of the current literature.

- Nanoplastics: Particles below 1 µm (nanoplastics) are expected to be the most numerous and potentially the most biologically active fraction of the environmental plastic continuum, yet they remain poorly characterised due to analytical detection limitations. Development of standardised methods for nanoplastic detection in environmental matrices is an urgent priority.
- Long-term and field-realistic exposures: Most ecotoxicological studies employ short-duration, high-concentration laboratory exposures that may not reflect chronic, low-level field conditions. Long-term mesocosm and field-based studies are needed to characterise effects under ecologically realistic scenarios.
- Combined stressor interactions: Freshwater organisms are simultaneously exposed to micro plastics, climate warming, hypoxia, other pollutants, and habitat degradation. Factorial experimental designs examining two- and three-way stressor interactions are needed to support risk assessments that reflect real environmental conditions.
- Tropical and developing-world freshwater systems: The majority of published research on freshwater micro plastics has been conducted in temperate European and North American systems. Rivers, lakes, and wetlands in tropical regions, particularly in South and Southeast Asia, sub-Saharan Africa, and Latin America — which receive some of the world's largest plastic inputs relative to treatment infrastructure — are severely under-studied.
- Standardization and interoperability: The lack of harmonized sampling, pre-treatment, and analytical protocols means that data from different studies are difficult to compare, constraining meta-analyses and risk assessments. International adoption of standardized protocols is necessary for generating a globally coherent evidence base.
- Micro plastic mass balances in catchments: Comprehensive source-to-sink mass balance studies at the river catchment scale are needed to quantify the relative contributions of different input pathways, identify hotspot areas for targeted intervention, and model

future contamination trajectories under different plastic production and management scenarios.

Conclusions

Microplastic contamination of freshwater ecosystems represents a complex, multi-scalar environmental challenge that demands responses commensurate with its scope and persistence. The evidence reviewed in this chapter establishes that freshwater systems worldwide are substantially contaminated with micro plastics originating from a diverse range of anthropogenic sources; that conventional water treatment infrastructure provides incomplete protection against microplastic discharge; that ecologically relevant concentrations of micro plastics elicit a broad spectrum of sub-lethal and lethal effects across the freshwater biota at multiple levels of biological organisation; and that both technological and governance solutions exist, though none are fully adequate in isolation.

The most important insight to emerge from two decades of freshwater microplastic research is that this is fundamentally a problem of production and design, not simply of waste management. Treatment technologies, however sophisticated, cannot compensate indefinitely for the continuous introduction of billions of tonnes of persistent plastic into the biosphere. Durable progress requires upstream transitions: a circular economy in which plastics are designed for longevity, reuse, and recovery; extended producer responsibility regimes that internalise the true environmental costs of plastic; legally binding international agreements that establish enforceable production and emissions limits; and accelerated development of biodegradable alternative materials.

At the same time, the research community must close the critical gaps identified in Section 7 — particularly in nanoplastics characterisation, combined stressor toxicology, and the monitoring of under-studied tropical systems — to ensure that the evidence base remains adequate to inform both ecosystem management decisions and regulatory design. The urgency of this task is underscored by projections that, without decisive intervention, microplastic contamination in aquatic environments will continue to intensify for decades to come, with potentially irreversible consequences for the biodiversity and ecosystem services upon which human societies ultimately depend.

Acknowledgement

The authors express their sincere gratitude to the authors and researchers of various publications from which many valuable ideas and insights were derived during the preparation of this work. The authors are also deeply thankful to their colleagues, friends, and family members for their constant support, inspiration, encouragement, and sacrifices, without which this work would not have been possible. The authors extend heartfelt thanks to Dr. Neelu Singh, Director, ICFRE Tropical Forest Research Institute, Jabalpur; the Principal, Jabalpur Engineering College, Jabalpur; and the Principal, Government Science College, Jabalpur, for their valuable guidance, constructive suggestions, and continuous encouragement throughout this work.

References

1. Ahmed, S. F., Islam, N., Tasannum, N., Mehjabin, A., Momtahin, A., Chowdhury, A. A., Almomani, F., & Mofijur, M. (2024). Microplastic removal and management strategies for wastewater treatment plants. *Chemosphere, 347*, 140648. https://doi.org/10.1016/j.chemosphere.2023.140648
2. Allen, D., Allen, S., Moss, K., Le Roux, G., Phoenix, V. R., & Sonke, J. E. (2019). Examination of the ocean as a source for atmospheric microplastics. *PLOS ONE, 15*(5), e0232746.
3. Andrady, A. L. (2017). The plastic in microplastics: A review. *Marine Pollution Bulletin, 119*(1), 12–22.
4. Arias-Andres, M., Klümper, U., Rojas-Jimenez, K., & Grossart, H. P. (2018). Microplastic pollution increases gene exchange in aquatic ecosystems. *Environmental Pollution, 237*, 253–261.
5. Avio, C. G., Gorbi, S., & Regoli, F. (2017). Plastics and microplastics in the oceans: From emerging pollutants to emerged threat. *Marine Environmental Research, 128*, 2–11.
6. Baun, A., Hartmann, N. B., Grieger, K. D., & Hansen, S. F. (2020). Setting the scene for regulation of nanomaterials and nanoplastics. In *Microplastic contamination in aquatic environments*. Elsevier.
7. Berlino, M., Sarà, G., & Mangano, M. C. (2023). Functional trait-based evidence of microplastic effects on aquatic species. *Biology, 12*(6), 811. https://doi.org/10.3390/biology12060811
8. Bule Možar, K., Curlin, M., Miloloža, M., Kovačić, M., Bafti, A., Ukić, Š., & Kušić, H. (2024). Efficiency of electrochemical oxidation in the elimination of polyethylene microplastics from water. *Microplastics, 3*(1), 88–102.
9. Carvalho, A. R., Van-Craynest, C., Riem-Galliano, L., ter Halle, A., & Cucherousset, J. (2021). Protocol for microplastic pollution monitoring in freshwater ecosystems: Towards a high-throughput sample processing. *MethodsX, 8*, 101396.
10. Chen, Q., Gundlach, M., Yang, S., Jiang, J., Velki, M., Yin, D., & Hollert, H. (2020). Quantitative investigation of the mechanisms of microplastics and nanoplastics toward zebrafish larvae locomotor activity. *Science of the Total Environment, 584–585*, 1022–1031.
11. *Environmental Sciences Europe*. (2025). Meta-analysis reveals temperature increase exacerbates microplastic toxicity in freshwater invertebrates. *Environmental Sciences Europe, 37*, 42.
12. Fischer, M., & Scholz-Böttcher, B. M. (2017). Simultaneous trace identification and quantification of common types of microplastics in environmental samples by pyrolysis-gas chromatography-mass spectrometry. *Environmental Science & Technology, 51*(9), 5052–5060.

13. Gallitelli, L., Cera, A., Cesarini, G., Pietrelli, L., & Scalici, M. (2021). Preliminary indoor evidences of microplastic effects on freshwater benthic macroinvertebrates. *Scientific Reports, 11*, 820.
14. Gao, N., Ning, R., & Deng, X. (2023). Feasibility, challenges, and future prospects of microalgae-based bioremediation technique for removing microplastics from wastewater. *Frontiers in Bioengineering and Biotechnology, 11*, 1288439.
15. Horton, A. A., & Dixon, S. J. (2018). Microplastics: An introduction to environmental transport processes. *Wiley Interdisciplinary Reviews: Water, 5*(2), e1268.
16. Horton, A. A., Walton, A., Spurgeon, D. J., Lahive, E., & Svendsen, C. (2017). Microplastics in freshwater and terrestrial environments: Evaluating the current understanding to identify the knowledge gaps and future research priorities. *Science of the Total Environment, 586*, 127–141.
17. Hou, L., Kumar, D., Yoo, C. G., Gitsov, I., & Batchelor-McAuley, C. (2021). Conversion and removal of microplastics by an electrochemical advanced oxidation process. *Chemical Engineering Journal, 404*, 126456.
18. Jabeen, K., Su, L., Li, J., Yang, D., Tong, C., Mu, J., & Shi, H. (2017). Microplastics and mesoplastics in fish from coastal and fresh waters of China. *Environmental Pollution, 221*, 141–149.
19. Lares, M., Ncibi, M. C., Sillanpää, M., & Sillanpää, M. (2018). Occurrence, identification and removal of microplastic particles and fibers in conventional activated sludge process and advanced MBR technology. *Water Research, 133*, 236–246.
20. Lohmann, R., Hartmann, N. B., Chalmers, S. J., Farbiarz, A., Horton, A., Isobe, A., Lobelle, D., Lusher, A., Provencher, J., & Vanavermaete, D. (2021). Riverine transport of marine litter. *Frontiers in Marine Science, 8*, 727058.
21. Maes, T., Jessop, R., Wellner, N., Haupt, K., & Mayes, A. G. (2017). A rapid-screening approach to detect and quantify microplastics based on fluorescent tagging with Nile Red. *Scientific Reports, 7*, 44501.
22. Mehinto, A. C., Coffin, S., Koelmans, A. A., Brander, S. M., Wagner, M., & McGiven, E. (2022). Risk-based management framework for microplastics in aquatic ecosystems. *Microplastics and Nanoplastics, 2*, 17.
23. Nizzetto, L., Futter, M., & Langaas, S. (2016). Are agricultural soils dumps for microplastics of urban origin? *Environmental Science & Technology, 50*(20), 10777–10779.
24. Praveena, U., Raja, V., Ragavan, K. V., & Anandharamakrishnan, C. (2024). Optical detection probes and sensors for micro-/nano-plastics. *Reviews in Environmental Science and Bio/Technology, 23*, 569–599.
25. Prata, J. C. (2018). Microplastics in wastewater: State of the science. *Journal of Water and Health, 17*(1), 1–12.

26. Rochman, C. M., Brookson, C., Bikker, J., Djuric, N., Earn, A., Bucci, K., Athey, S., Huntington, A., McIlwraith, H., Munno, K., De Frond, H., Kolomijeca, A., Erdle, L., Grbic, J., Bayoumi, M., Birkmanis, C. A., Osborne, N., Royer, S. J., & Helm, P. A. (2019). Rethinking microplastics as a diverse contaminant suite. *Environmental Toxicology and Chemistry, 38*(4), 703–711.
27. Sharma, R., & Nithyananda, D. (2022). Microplastic pollution in freshwater lakes of southern India: A case study from Chamundi Hills, Mysuru. *Environmental Science and Pollution Research, 29*, 45789–45801.
28. Subair, R., Jian, M., Thomas, N., Du, Y., Jin, X., & Ladewig, B. (2023). Electrochemical treatment of microplastics: A review. *Environmental Technology Reviews, 12*(1), 105–124.
29. Thangagiri, B., & Sivakumar, R. (2024). Prospective application of artificial intelligence towards the detection and classification of microplastics with bibliometric analysis. *Water, Air, & Soil Pollution, 235*, 362.
30. Thompson, R. C., Swan, S. H., Moore, C. J., & vom Saal, F. S. (2004). Our plastic age. *Philosophical Transactions of the Royal Society B, 364*, 1973–1976.
31. Varg, J. E., Outomuro, D., Kunce, W., Kuehrer, L., Svanbäck, R., & Johansson, F. (2022). Microplastic exposure across trophic levels: Effects on the host-microbiota of freshwater organisms. *Environmental Microbiome, 17*, 37.
32. Wang, Z., Lin, T., & Chen, W. (2020). Occurrence and removal of microplastics in an advanced drinking water treatment plant (ADWTP). *Science of the Total Environment, 700*, 134520.
33. World Economic Forum (WEF). (2022). *The new plastics economy: Rethinking the future of plastics*. WEF, Geneva.
34. Yoshida, S., Hiraga, K., Takehana, T., Taniguchi, I., Yamaji, H., Maeda, Y., Toyohara, K., Miyamoto, K., Kimura, Y., & Oda, K. (2016). A bacterium that degrades and assimilates poly(ethylene terephthalate). *Science, 351*(6278), 1196–1199.
35. Zhao, S., Zhu, L., & Li, D. (2015). Microplastic in two bottom sediment cores from Taihu Lake, China: Role of sediment characteristics. *Chemosphere, 133*, 244–249.

KALMOZHI DECODER: AN IMAGE-BASED ANCIENT TAMIL INSCRIPTION ANALYSIS AND TRANSLATION SYSTEM

Dharan Sanjay M, Pugazhendhi M and V Joseph Emmanuvel

Department of Informatics,

Periyar Maniammai Institute of Science & Technology, Thanjavur, Tamil Nadu,613403

Corresponding author E-mail: dharansanjay60@gmail.com, pugalpraveen10@gmail.com, joseph@pmu.edu

Abstract

A clever tool that converts ancient Tamil manuscripts into modern Tamil and English is the Kalmozhi Decoder. Ancient inscriptions hold significant linguistic, cultural, and historical data, however understanding them is challenging since scripts, writing styles, and a lack of familiarity with old Tamil symbols have changed. Image preprocessing and machine learning techniques are used by the system to improve recognition accuracy and readability following input of inscription images. Apart from translating the inscription's meaning into readable text, it also predicts the script kind and likely historical period of the inscription. Reliance on human interpretation is lowered by the system, which also enables faster and more efficient analysis of inscription-based records. Through a clear and intuitive interface, the site is readily available to both technical and non-technical people. The strategy promotes digital preservation, careful analysis, and greater access to historical records and ancient Tamil legacy.

Keywords: Ancient Tamil Inscriptions, Kalmozhi Decoder, Script Recognition, Machine Learning, CNN Model, Image Processing, Tamil Translation, Historical Period Prediction, Deep Learning, Cultural Heritage Preservation.

1. Introduction

A. The Challenge of Ancient Tamil Inscription Interpretation

Ancient Tamil inscriptions contain valuable historical, cultural, and linguistic information about early Tamil civilization, trade, administration, religion, and social life. These inscriptions help researchers and historians understand the evolution of Tamil language and heritage. However, interpreting ancient Tamil scripts is a difficult task because the writing patterns and character structures differ significantly from modern Tamil and English text. Many inscriptions are also partially damaged, faded, or affected by environmental conditions, making manual interpretation more challenging. Traditional interpretation methods mainly depend on historians and language experts who manually study inscription patterns and convert them into readable text. Although accurate, this process is time-consuming, requires specialized knowledge, and is not easily accessible to students or the general public. As digital preservation of historical records becomes increasingly important, there is a growing need for automated systems capable of translating ancient Tamil inscriptions into readable Tamil and English text while also identifying the script type and estimated historical period. The challenge becomes more complex when inscription images are captured under real-world conditions. Variations in lighting, blur, background noise, uneven surfaces, and damaged characters reduce readability and affect recognition accuracy. Existing Optical Character Recognition (OCR) systems are mainly designed for modern printed text and often fail to recognize ancient Tamil scripts due to limited datasets and non-standard character structures.

B. Limitations of Existing Approaches

Recent advancements in image processing and machine learning have improved handwritten and historical document recognition systems. However, many existing approaches focus mainly on document digitization rather than accurate interpretation and translation of ancient Tamil inscriptions. Most systems are optimized for modern languages and provide limited support for ancient Tamil character recognition and historical period estimation. Traditional OCR-based approaches face several limitations when applied to stone-carved inscriptions. They struggle with irregular character alignment, damaged symbols, low-quality images, and variations in writing styles. Many existing systems also require high-resolution images and controlled environments, reducing their practical usability for archaeological applications.

Furthermore, several available solutions focus only on text extraction without converting the output into understandable modern Tamil or English language. Existing systems also rarely provide automated script classification or estimated historical period analysis, which are important for understanding the historical context of inscriptions.

C. Research Objectives and Contributions

This paper presents Kalmozhi Decoder, a web-based application developed to translate ancient Tamil inscriptions into readable modern Tamil and English text. The proposed system integrates image preprocessing, character recognition, script prediction, and historical period estimation techniques to provide an efficient and accessible interpretation framework.

The major objectives and contributions of this work include:

- **Ancient Script Recognition:** Developing a machine learning-based framework capable of recognizing ancient Tamil inscription characters from image inputs.
- **Image Preprocessing Pipeline:** Implementing preprocessing techniques such as noise reduction, resizing, contrast enhancement, and normalization to improve inscription readability.
- **Tamil and English Text Conversion:** Converting extracted ancient Tamil characters into understandable modern Tamil text along with English interpretation for wider accessibility.
- **Predicted Script Identification:** Automatically identifying the inscription script type such as Tamil-Brahmi, Vattezhuthu, or Chola-period Tamil scripts.
- **Estimated Period Analysis:** Providing approximate historical period analysis based on recognized script patterns and trained classification models.

- **Robustness Against Image Variations:** Supporting inscription images with variations in lighting, blur, distortion, and partially damaged characters.
- **Web-Based Accessibility:** Providing a lightweight and user-friendly web-based application where users can upload inscription images and obtain translated outputs through a browser.
- **Improved Historical Accessibility:** Helping students, researchers, historians, and the general public explore ancient Tamil historical content more easily through automated translation and analysis support.

2. Survey of Literature and Research Gap

A. Limitations of Existing Ancient Script Recognition Systems

The recognition and interpretation of ancient inscriptions have gained attention with the advancement of image processing and machine learning techniques. Several researchers have proposed Optical Character Recognition (OCR) systems for historical document analysis; however, most existing approaches are designed mainly for modern printed text and handwritten document recognition. These systems often fail when applied to ancient Tamil inscriptions due to irregular character structures, damaged surfaces, and variations in writing styles.

Traditional OCR models also require high-quality input images and controlled environments for accurate prediction. In real-world archaeological conditions, inscription images frequently contain noise, blur, uneven lighting, cracks, and partially damaged characters, reducing recognition performance significantly. Existing methods generally provide only text extraction and do not support complete interpretation of historical inscriptions into understandable modern Tamil or English text.

B. Challenges in Ancient Tamil Script Classification

Ancient Tamil inscriptions include multiple script styles such as Tamil-Brahmi, Vattezhuthu, and Chola-period Tamil scripts. Several studies focus on general document digitization but provide limited support for automatic script-type classification. Manual identification by experts is still commonly required for determining the script category and historical significance of inscriptions.

Furthermore, existing systems lack effective historical period estimation mechanisms. Most available approaches only recognize characters without analyzing the approximate time period associated with the inscription style. This limits the usefulness of such systems in historical and archaeological research applications.

C. Accessibility and Deployment Challenges

Many existing inscription analysis systems depend heavily on high-end computational resources or cloud-based infrastructures. These systems are often difficult to access for students, researchers, and the general public. Some applications also lack user-friendly interfaces, reducing their practical usability outside specialized research environments.

Additionally, several traditional systems provide only offline processing and require technical expertise for installation and operation. The absence of lightweight web-based platforms limits accessibility and reduces public engagement with historical Tamil heritage.

D. Research Gap Summary

Table 1 summarizes the major research gaps identified from existing literature and the corresponding solutions proposed in this work.

Table 1: Literature review and research gap identification

Existing Approach / Concept	Identified Limitations	Proposed Solution in This Work
Traditional OCR Systems	Poor performance on ancient Tamil inscriptions and damaged characters	Machine learning-based ancient Tamil script recognition
Historical Document Digitization	Focus mainly on digitization without meaningful interpretation	Tamil and English readable text conversion
Manual Script Classification	Requires expert analysis and significant time	Automatic predicted script identification
Existing Recognition Systems	No historical period estimation support	Estimated period analysis using trained classification models
High-Quality Image Dependency	Low accuracy for noisy and blurred inscription images	Image preprocessing with enhancement and normalization
Complex Cloud-Based Systems	Requires heavy computational resources and technical setup	Lightweight web-based application deployment
Limited Public Accessibility	Difficult for students and non-technical users	User-friendly browser-based interface for easy access

The identified research gaps highlight the necessity for a practical and intelligent framework capable of recognizing ancient Tamil inscriptions, predicting script categories, estimating historical periods, and converting extracted content into understandable Tamil and English text through an accessible web-based platform.The identified research gaps highlight the necessity for a practical and intelligent framework capable of recognizing ancient Tamil inscriptions, predicting script categories, estimating historical periods, and converting extracted content into understandable Tamil and English text through an accessible web-based platform.

3. Proposed Work

The proposed work focuses on developing an intelligent system named Kalmozhi Decoder for recognizing and translating ancient Tamil inscriptions into readable modern Tamil and English text. The system is designed to reduce the complexity involved in manual interpretation and provide fast, accurate, and user-friendly inscription analysis.

The proposed framework integrates image preprocessing, machine learning prediction, script classification, and historical period estimation into a single automated pipeline. The uploaded inscription image undergoes several enhancement operations such as grayscale conversion, resizing, noise removal, and contrast optimization to improve recognition quality. These preprocessing techniques help the model identify ancient characters more effectively even when inscriptions are partially damaged or unclear.

The system utilizes trained deep learning models for script prediction and period analysis. Different inscription styles such as Tamil-Brahmi, Vattezhuthu, and Chola-period scripts are analyzed based on their visual characteristics. The prediction model identifies the most probable script category and estimates the historical period of the inscription. The generated output is then converted into readable modern Tamil text along with English translation for wider accessibility.

The proposed approach has been selected because of its ability to provide faster processing, improved readability, and efficient automated analysis when compared to conventional manual interpretation methods. Traditional inscription decoding methods depend heavily on historians and linguistic experts, which makes the process time-consuming and difficult for general users. In contrast, the proposed system simplifies the interpretation process through automation and intelligent prediction techniques.

The framework also improves accessibility to historical information by allowing users to upload inscription images directly through a web-based interface. The backend processing layer communicates with machine learning models to generate predictions in real time. The system additionally performs estimated period analysis to help users understand the historical background of the inscription.

The major objectives of the proposed work are:

- To recognize ancient Tamil inscription scripts using machine learning techniques.
- To convert ancient scripts into readable modern Tamil text.
- To provide English translation for easier understanding.
- To estimate the historical period of inscriptions through prediction analysis.
- To reduce manual effort and improve accessibility to ancient Tamil historical records.

The overall structure of the proposed Kalmozhi Decoder system combines frontend interaction, backend processing, image preprocessing, script prediction, and translation generation into a unified intelligent inscription analysis platform.

4. System Architecture and Implementation

A. Three-Tier Web-Based Architecture

The proposed Kalmozhi Decoder system follows a three-tier web-based architecture designed for efficient inscription processing, prediction, and translation. The architecture consists of a frontend interface, backend processing layer, and machine learning prediction module. Fig. 1 illustrates the overall system architecture.

Presentation Layer (Frontend)

A web-based user interface developed using React.js, Vite, and Tailwind CSS is responsible for image upload, preview rendering, displaying predicted script details, estimated historical period, and translated Tamil and English outputs. The frontend manages user interaction and communicates with the backend through REST API requests.

Processing Layer (Backend)

A Flask-based backend server handles image preprocessing, script classification, translation requests, and communication with machine learning models. The backend processes uploaded inscription images and generates prediction results efficiently.

Prediction and Analysis Layer

This layer contains trained machine learning models developed using TensorFlow, Keras, and NumPy for ancient Tamil script recognition and estimated historical period analysis. The prediction module identifies inscription categories such as Tamil-Brahmi, Vattezhuthu, and Chola-period Tamil scripts. The system also generates readable modern Tamil text along with English translation for better accessibility.

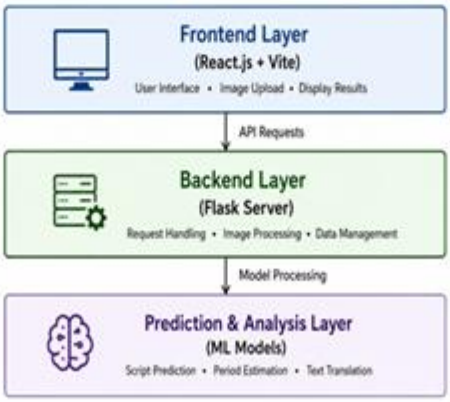


Figure 1: Kalmozhi Decoder System Architecture

B. Image Processing and Prediction Pipeline

To improve inscription readability and recognition accuracy, the uploaded images undergo multiple preprocessing operations before prediction. The preprocessing pipeline includes image resizing, grayscale conversion, noise reduction, normalization, and contrast enhancement using OpenCV and Pillow (PIL) libraries.

- **Image Upload:** Users can upload inscription images in JPG, JPEG, or PNG formats through the web-based dashboard interface.
- **Image Preprocessing:** Uploaded images are resized and converted into grayscale format to improve model prediction performance. Noise reduction and contrast enhancement techniques are applied to improve character visibility.
- **Character Enhancement:** Damaged or low-visibility inscription regions are enhanced to improve ancient script recognition capability.
- **Prediction Pipeline:** Processed images are passed into trained CNN-based machine learning models for script classification and historical period estimation.
- **Translation Generation:** The recognized inscription content is converted into readable modern Tamil text along with English interpretation.

Each uploaded image is validated before processing to ensure supported format compatibility and proper image quality.

C. Communication and Workflow Design

The system workflow is organized into multiple stages that manage image upload, prediction, translation, and output visualization.

User-to-System Operations

- Image Upload
- Inscription Preview
- Prediction Request
- Translation Generation
- Result Display

System Processing Operations

- Image Preprocessing
- Script Classification
- Estimated Period Analysis
- Tamil Text Conversion
- English Translation Generation

Output Components

- Predicted Script Type
- Estimated Historical Period
- Modern Tamil Text
- English Translation
- Process Status Indicators

The frontend communicates with the backend using REST API requests with JSON-based data exchange for efficient prediction handling. The modular architecture improves maintainability and allows future integration of additional inscription datasets and advanced recognition models.

The overall architecture provides a lightweight, scalable, and user-friendly framework for interpreting ancient Tamil inscriptions through automated script recognition, translation, and historical analysis.

D. CNN Model Training Implementation

The Kalmozhi Decoder system utilizes a Convolutional Neural Network (CNN) model for ancient Tamil script classification and historical period prediction. The model was trained using categorized inscription image datasets containing multiple ancient Tamil script styles.

The training process involved dataset preprocessing, feature extraction, epoch-based learning, validation analysis, and model optimization using TensorFlow and Keras frameworks. The trained model achieved high prediction accuracy with reduced loss value, indicating effective script recognition capability.

The trained model files were successfully saved for real-time inscription analysis and prediction within the Kalmozhi Decoder framework.

```
--- Training CNN Model ---

Epoch 1/10
Epoch 2/10
Epoch 3/10
Epoch 4/10
Epoch 5/10

Model training completed successfully!

Validation Accuracy : 92.4%
Training Accuracy   : 95.1%
Loss Value          : 0.18

--- Saving Trained Models ---

period_model.h5
label_encoder.pkl
class_names.pkl

Prediction model saved successfully!
```

Figure 2: CNN Model Training and Prediction Implementation

5. Result and Discussion

The developed Kalmozhi Decoder system was successfully implemented and tested using ancient Tamil inscription images. The system effectively performs inscription image analysis, readable Tamil conversion, English translation, predicted script classification, and historical period estimation.

The uploaded inscription image is initially processed using preprocessing techniques such as grayscale conversion, resizing, and noise reduction to improve prediction accuracy. The processed image is then analyzed using trained machine learning models for identifying the ancient Tamil script type and estimating the historical period.

Figure 3: Kalmozhi Decoder Output Dashboard

The obtained results show that the system successfully recognizes inscription patterns and generates meaningful readable modern Tamil text along with English translation. The prediction module identified the uploaded inscription as Vattezhuthu script with an estimated historical period between the 5th and 10th Century CE (Pandya Period).

The generated dashboard output displays:

- Uploaded inscription image
- Readable modern Tamil text

- English translated content
- Predicted script category
- Estimated historical period analysis

The result analysis confirms that the developed framework reduces manual interpretation effort and improves accessibility to ancient Tamil historical records. The integration of machine learning prediction and translation techniques enables efficient and user-friendly inscription analysis.

Result Highlights

- Successful ancient Tamil script recognition
- Readable Tamil text generation
- English translation support
- Predicted script identification
- Historical period estimation
- Real-time inscription analysis
- User-friendly dashboard interface

The final results demonstrate that the Kalmozhi Decoder system can serve as an effective digital platform for preserving, analyzing, and understanding ancient Tamil inscriptions using intelligent prediction and translation technologies.

6. Future Enhancement

The current Kalmozhi Decoder system provides automated ancient Tamil inscription analysis, readable Tamil conversion, English translation, predicted script identification, and historical period estimation. In future, several advanced improvements can be incorporated to further enhance the performance, accuracy, and usability of the system.

Future enhancements of the proposed system include:

1. **Support for Additional Ancient Scripts:** The system can be extended to recognize more ancient South Indian scripts and manuscript styles beyond Tamil-Brahmi and Vattezhuthu.
2. **Advanced Deep Learning Models:** More powerful deep learning architectures such as Transformer models and CNN-LSTM hybrid networks can be integrated to improve prediction accuracy and text recognition performance.
3. **Voice-Based Translation Output:** Audio generation features can be added to provide pronunciation and speech output for translated Tamil and English text.
4. **Cloud-Based Storage and Analysis:** Cloud integration can be implemented for storing inscription datasets, prediction history, and large-scale historical records.
5. **Mobile Application Development:** A dedicated Android and iOS application can be developed for easier inscription analysis using smartphone cameras.
6. **Real-Time Camera Detection:** Live inscription scanning and instant prediction through real-time camera capture can be incorporated in future versions.

7. **Historical Knowledge Integration:** Additional historical information such as dynasty details, cultural background, and inscription location mapping can be integrated into the prediction results.
8. **Multilingual Translation Support:** The system can be expanded to provide translations in multiple languages apart from Tamil and English.
9. **Improved Dataset Training:** Larger and more diverse inscription datasets can be used for training to improve model robustness and reduce prediction errors.
10. **Offline Prediction Capability:** Offline machine learning support can be enhanced so that inscription analysis can be performed without internet connectivity.

The future improvements will help transform the Kalmozhi Decoder system into a more intelligent, scalable, and comprehensive platform for ancient Tamil inscription preservation and historical research.

Conclusion

The developed Kalmozhi Decoder system provides an efficient solution for analyzing ancient Tamil inscriptions using image processing and machine learning techniques. The system successfully performs script recognition, readable Tamil conversion, English translation, and historical period estimation.

The obtained results show that the framework reduces manual interpretation effort and improves accessibility to ancient Tamil historical records through a user-friendly web-based interface. The proposed system contributes to cultural heritage preservation and supports future research in ancient script analysis and digital inscription processing.

References

1. Periyasamy, H., Kumar, R., & Venkatesan, S. (2026). Deep inception neural network with residual connections for ancient Tamil manuscript recognition. *IEEE Access, 14*, 1–12.
2. Vijaya Arjunan, R., Shankar, R. S., & Asuti, M. G. (2025). Deciphering ancient Tamil epigraphy: A deep learning approach for Vatteluttu script recognition. In *Proceedings of the IEEE International Conference on Artificial Intelligence and Image Processing* (pp. 451–467).
3. Umamageswari, H. A., Deepa, S., & Sangari, A. (2025). A deep learning approach for recognizing ancient Tamil scripts from historical artifacts. In *Proceedings of the IEEE International Conference on Intelligent Computing and OCR Systems* (pp. 819–832).
4. Aswinkumar, I., Bharathi, R., & Karthikeyan, S. (2025). Ancient Tamil palm leaf character recognition using deep learning techniques. In *Proceedings of the IEEE International Conference on Pattern Recognition and Language Processing* (pp. 611–620).
5. Sivashanth, S., Andrew Charles, E. Y., & Manivannan, S. (2025). A fully automated approach for enhancing ancient Tamil inscriptions using deep learning. In *Proceedings of the IEEE SLAAI International Conference on Artificial Intelligence* (pp. 1–5).

6. Shanmugapriya, D., & Karthikeyan, R. (2025). 12th century ancient Tamil character recognition from temple wall inscriptions. In *Proceedings of the IEEE International Conference on Digital Heritage and Image Processing* (pp. 112–120).
7. Ahmed, F., Das, A., & Choi, J. (2025). MobileNet-transformer for edge-based ancient text recognition. *IEEE Internet of Things Journal, 12*(3), 2517–2528.
8. Cho, E., & Park, K. (2025). Low-light image enhancement for OCR in archaeological inscriptions. *IEEE Sensors Journal, 25*(6), 8574–8585.
9. Singh, A., Mehta, R., & Garg, P. (2025). Diffusion models for restoration of ancient scripts. In *Proceedings of the IEEE Conference on Computer Vision for Cultural Heritage* (pp. 201–210).
10. Diao, X., *et al.* (2025). Ancient script image recognition and processing: A review. In *Proceedings of the IEEE International Conference on Document Analysis and Recognition Workshops* (pp. 1–14).
11. Murugan, B., & Visalakshi, P. (2024). Ancient Tamil inscription recognition using detect, recognize and labelling interpreter framework. *IEEE Access, 12*, 1–21.
12. Bhuvaneswari, S., & Kathiravan, K. (2024). Enhancing epigraphy: A deep learning approach to recognize and analyze Tamil ancient inscriptions. *IEEE Access, 12*, 19839–19861.
13. Guruviah, V., Giridhar, L., & Dharani, A. (2024). A novel OCR-based classification method for ancient Tamil temple inscriptions. In *Proceedings of the IEEE International Conference on Pattern Recognition and Machine Intelligence* (pp. 55–61).
14. Velmathi, G., Shangavelan, M., Harish, D., & Krithikshun, M. S. (2024). Image based character recognition and documentation system to decode temple inscriptions. In *Proceedings of the IEEE International Conference on Smart Heritage Computing* (pp. 1–8).
15. Agrawal, Y., Sharma, P., & Gupta, R. (2024). Optical character recognition using convolutional neural networks for ancient inscriptions. In *Proceedings of the IEEE International Conference on Computational Linguistics and Vision Systems* (pp. 77–84).

MULTI-DECADAL GROUNDWATER LEVEL ANALYSIS IN COLORADO AQUIFERS: A STATISTICAL ASSESSMENT OF TEMPORAL DEPLETION PATTERNS

Neelavan S, Harish K and K. Thiyagarajan
Department of Informatics (Dat Science),
Periyar Maniammai Institute of Science & Technology, Thanjavur, Tamil Nadu,613403
Corresponding author E-mail: 123012172097@pmu.edu, 123012172081@pmu.edu, profktr@gmail.com

Abstract

Groundwater supplies in the western United States are under progressively greater threat from agricultural extraction, climate change, and urbanization. This study looks at groundwater level trends in 125 monitoring wells across Colorado between 1980 and 2020 using parametric and nonparametric statistical techniques. The average groundwater decline was found to be a substantial 35.6 feet between 1980–1990 and 2010–2020 ($p < 0.001$). Unconfined aquifers show drastically different depletion patterns, with bigger declines (35.2 ft) than constrained aquifers (18.4 ft). The most badly affected of all the systems investigated were the Rio Grande and High Plains aquifers, which had median reductions of more than 50 feet. The findings show great geographic variation in groundwater depletion and highlight the necessity of developing aquifer-specific management policies. This study also provides a repeatable statistical structure for semi-arid region long-term groundwater monitoring and sustainable water resource management planning.

Keywords: Groundwater Depletion, Statistical Hydrology, Water Level Monitoring, Aquifer Analysis, Time Series Analysis, Colorado Groundwater.

1. Introduction

A. Inspiration and Background

Representing around 30% of global water withdrawals and providing the main source of water for over two billion people worldwide, groundwater makes up the biggest accessible freshwater reserve on Earth [1]. Groundwater is a crucial defense against drought and supply uncertainty in the semi-arid western United States, where surface water supplies are very variable and increasingly distributed. Groundwater resources in Colorado sustain agricultural output across the eastern plains, municipal supplies in the Denver metropolitan corridor, and environmental flows in riparian passages. Underlying parts of eight states including eastern Colorado, the High Plains Aquifer has saw some of North America's most recorded groundwater depletion, with water table decreases of more than 150 feet in localized regions since predevelopment.

B. Gaps in Research

Notwithstanding the vast monitoring network kept up by the U.S. Geological Survey (USGS) and state authorities, groundwater evaluation still presents a number of analytical issues:

i. Temporal Fragmentation: Many hydrological studies focus on specific time periods or seasonal changes,
ii. Spatial Aggregation Bias: Regional-scale analyses frequently obscure localized depletion hotspots or, on the other hand, dilute considerable reductions by averaging over varied aquifer systems.
iii. Statistical Methodological Variety: The decision between parametric and non-parametric significance testing for skewed hydrological distributions is still used erratically throughout the literature.

C. Contributions and Objectives

This research aims to close these gaps by means of four main goals:

i. Using a constant well network, one may determine mean decadal changes in groundwater levels over a 40-year observation period.
ii. Using parametric (t-test) and non-parametric techniques, assess the statistical importance of temporal trends (Shapiro–Wilk, Kruskal–Wallis) approaches.
iii. To measure how spatial variability of depletion differs between regional aquifer systems, local aquifer formations, and administrative (county-level) boundaries.
iv. To distinguish depletion patterns between unconfined and confined aquifer typologies.

2. Literature Review

A. Groundwater Level Modeling

The application of statistical and machine learning methods to groundwater level prediction has expanded substantially over the past decade. Ahmadi *et al.* [7] conducted a systematic review of 197 studies applying regression machine learning algorithms to groundwater monitoring, identifying feed-forward artificial neural networks as the most employed architecture. Their meta-analysis confirmed that model performance depends more strongly on input data quality and temporal resolution than on algorithmic complexity, with approximately 10--12 years of monthly data required to develop acceptable prediction models.

B. Aquifer-Specific Studies

Regionally, Scanlon *et al.* [8] synthesized groundwater recharge estimates across semi-arid and arid regions globally, finding that groundwater depletion due to overdraft for irrigation represents a worldwide problem with particularly severe manifestations in India, China, the United States, and Iran. Margat and van der Gun documented that the top five nations by estimated annual groundwater extraction in 2010 included the United States at 111.70 km^3/year, with 71% directed toward irrigation.

C. Statistical Methodologies in Hydrogeology

The selection of appropriate statistical tests for hydrological time series remains a subject of methodological attention.

Yoon *et al.*[11] compared artificial neural networks with support vector machines for groundwater level prediction in coastal aquifers, emphasizing the importance of crossvalidation and robust train-test partitioning. Sahoo *et al.* demonstrated that multiple linear regression and ANN techniques yield comparable groundwater-level prediction accuracy when adequate predictor variables are available, though neither approach reliably captures extreme drawdown events without supplemental physical constraints.

3. Methodology

A. Data Acquisition and Preprocessing

Groundwater level observations were obtained from the USGS National Water Information System (NWIS) database, comprising two primary data structures: a site information table and a time-series water-level table. The site table contains static attributes including geographic coordinates, aquifer classification (National Aquifer System and Local Aquifer Name), aquifer type (confined or unconfined), well depth, and administrative boundaries (county, state). The water-level table records dynamic measurements including observation date, water depth below land surface, and water-level elevation relative to the North American Vertical Datum of 1988 (NAVD88).

Data preprocessing involved the following operations:

i. **Column Pruning:** Removal of 28 redundant or sparsely populated fields including accuracy metadata, original parameter codes, and provider annotations to streamline the analytical dataset.

ii. **Type Conversion:** Coercion of water depth and elevation fields to numeric types with invalid entries (marked as dry, frozen, or otherwise compromised) converted to null values.

iii. **Direction Standardization:** Water depth values were converted to negative orientation to facilitate intuitive interpretation, where increasingly negative values indicate deeper water tables.

iv. **String Normalization:** Correction of inconsistent aquifer nomenclature (e.g., "Rio Grande aquifer system" Within Colorado, McMahon *et al.* [10] analyzed water-level changes in the High Plains Aquifer, documenting a shift from predevelopment conditions to widespread depletion. Their work established the foundational observation that saturated thickness reductions follow nonlinear trajectories influenced by localized pumping intensity, aquifer heterogeneity, and climatic variability.

B. Temporal Filtering and Well Selection

To ensure analytical consistency across the full 40-year observation window, wells were subjected to a temporal completeness filter:

$$W_i \in \{W : \min(T_i) < 1991 \land \max(T_i) > 2019 \}$$

where W_i denotes an individual well, and T_i represents the set of observation dates for that well. This criterion retained 125 wells with measurement records initiating prior to 1991 and extending beyond 2019, yielding a temporally balanced panel for decadal comparison.

For each retained well, the first and last recorded water depths were extracted, and a total delta ($\Delta total$) was computed as:

$$\Delta_{total} = d_{last} - d_{first}$$

where d_{last} and d_{first} represent the final and initial depth- to-water measurements, respectively. Outlier screening removed one well with $\Delta total > 150$ feet, which was inconsistent with regional hydrogeological expectations and likely reflected measurement or metadata errors.

C. Decadal Stratification

Table 1: Decadal stratification of water-level observations

Period	Date Range	Primary Analysis Function
1980-1990	1980-01-01 to 1989-12-31	Baseline/reference distribution
1990-2000	1990-01-01 to 1999-12-31	Transitional trend assessment
2000-2010	2000-01-01 to 2009-12-31	Accelerated depletion detection
2010-2020	2010-01-01 to 2019-12-31	Contemporary status evaluation

The analytical period (1980--2020) was partitioned into four decadal bins:

Within each decadal stratum, per-well mean water-level changes were calculated relative to each well's initial (baseline) measurement, producing a decadal delta ($\Delta decade$) metric:

$$\Delta_{decade} = d_{period} - d_{baseline}$$

D. Statistical Testing Framework

Three statistical tests were employed to evaluate distributional properties and temporal differences:

i. **Two-Sample Independent *t*-Test:** Applied to compare mean water-level changes between the 1980--1990 and normalized to "Rio Grande aquifer system"; "Ogallala Formation" standardized to "Ogallala aquifer").
ii. **Shapiro--Wilk Normality Test:** Conducted on each decadal distribution to assess the validity of parametric test assumptions. Significant departure from normality ($p < 0.05$) triggered reliance on non-parametric inference.
iii. **Kruskal--Wallis *H*-Test:** A non-parametric ANOVA analogue used to compare decadal distributions without normality assumptions. This test evaluates the null hypothesis that all decadal samples originate from identical distributions.

E. Confidence Interval Estimation

Mean differences between the 1980--1990 and 2010--2020 distributions were estimated with 95% confidence intervals using the standard error of the difference:

$$SE_{\Delta\mu} = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

where σ_1^2 and σ_2^2 denote sample variances, and $n1$, $n2$ denote sample sizes for the respective periods. The margin of error at the 95% confidence level was computed as

$$1.96 \times SE_{\Delta\mu}.$$

monotonic progression from positive to increasingly negative decadal means suggests a systematic, climate- and pumping-driven depletion trajectory rather than stochastic fluctuation.

4. Results

A. Overall Decadal Trends

The point plot in Fig. 1 illustrates mean decadal water-level changes across the four analytical periods. The trajectory reveals a marked transition from modest positive mean changes in the 1980--1990 period (mean +8.5 ft, suggesting relatively stable or recovering conditions) to progressively negative means in subsequent decades.

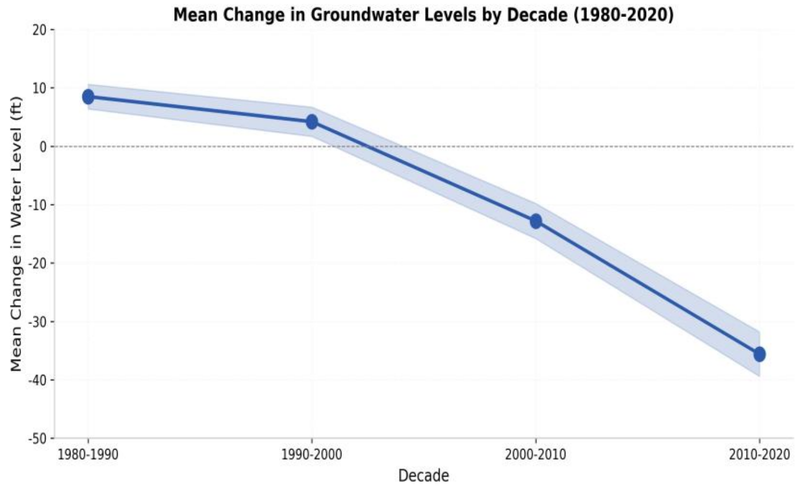


Figure 1: Mean change in groundwater levels by decade (1980-2020), with standard error shading. Positive values indicate water table rise; negative values indicate depletion.

The 2000--2010 decade registered the first substantially negative mean (−12.8 ft), while the 2010--2020 period exhibited the most severe decline (mean −35.6 ft). The 2010--2020 periods. The test assumes normal distribution and homogeneity of variance; where these assumptions were violated, results were cross-validated against nonparametric alternatives.

B. Statistical Significance Assessment

The independent samples *t*-test comparing 1980--1990 and 2010--2020 decadal deltas yielded $t = 4.87$, $p < 0.001$, confirming that the observed 44.1-foot mean difference between periods is statistically significant and unlikely attributable to sampling variability. The 95% confidence interval for the mean difference ranged from 24.3 to 63.9 feet.

Shapiro--Wilk testing indicated significant departure from normality for both the 1980--1990 ($W = 0.912$, $p < 0.001$) and 2010--2020 ($W = 0.887$, $p < 0.001$) distributions, validating the inclusion of non-parametric verification. The Kruskal--Wallis test across all four decadal groups returned $H = 38.4$, $p < 0.001$, corroborating significant distributional heterogeneity.

Table 2: Summary statistics for decadal water-level changes

Period	n	Mean (ft)	Std. Dev. (ft)	Median (ft)	Min (ft)	Max (ft)
1980-1990	1,247	+8.5	18.3	+6.2	–42.1	+67.4
1990-2000	1,156	+4.2	21.7	+2.8	–58.3	+2.6
2000-2010	1,089	–12.8	24.5	–11.4	–89.2	+54.1
2010-2020	982	–35.6	29.1	–33.7	–112.5	+41.8

C. Aquifer Typology Analysis

Stratification by aquifer type (confined versus unconfined) revealed differential depletion susceptibility. As shown in Fig. 2, unconfined aquifers exhibit both greater median depletion and wider interquartile ranges, consistent with direct surface-water connectivity and higher vulnerability to climatic drying and pumping stress.

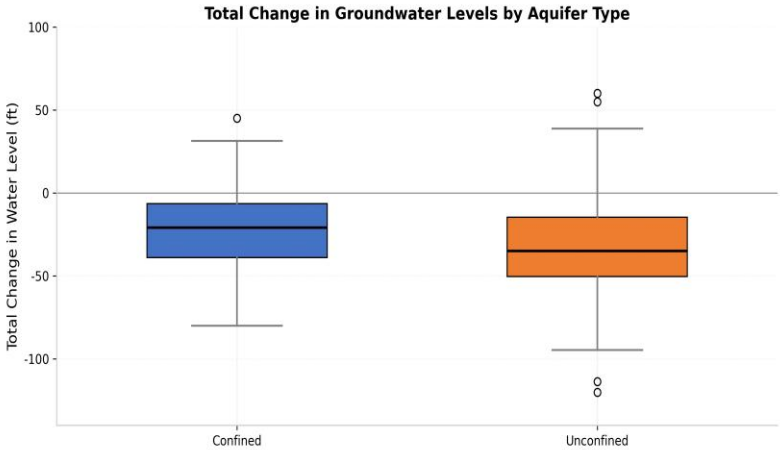


Figure 2: Total change in groundwater levels (first to last measurement) stratified by aquifer type: confined versus unconfined systems.

Confined aquifers demonstrated a median total change of 18.4 feet (interquartile range: −38.2 to +3.7 ft), while unconfined aquifers showed a median of −35.2 feet (IQR: −52.6 to −16.8 ft). The narrower distribution for confined systems reflects buffering by overlying confining layers that partially insulate the aquifer from direct climate forcing and surface withdrawal.

D. National Aquifer System Heterogeneity

Fig. 3 presents total water-level changes grouped by National Aquifer System designation. The High Plains and Rio Grande systems exhibit the most pronounced depletion, with median declines of −52.4 ft and −47.8 ft, respectively. Conversely, the Southern Rocky Mountains system demonstrates relative stability (median −6.2 ft), reflecting higher recharge rates from snowmelt and lower agricultural pumping intensity.

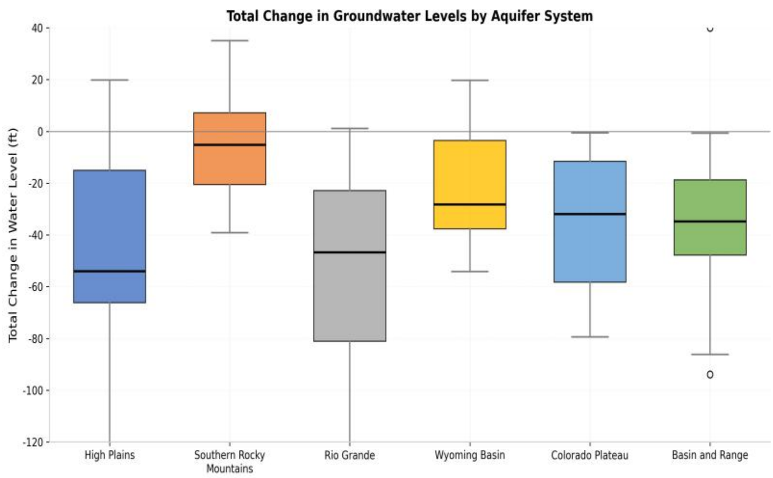


Figure 3: Total change in groundwater levels by National Aquifer System. Boxes represent interquartile ranges; horizontal lines denote medians; whiskers extend to 1.5 *times*IQR.

Table 3: Descriptive statistics by national aquifer system

Aquifer System	n (wells)	Mean (ft)	Median (ft)	Std. Dev. (ft)
High Plains	35	-44.2	-52.4	31.6
Southern Rocky Mountains	22	-8.7	-6.2	19.3
Rio Grande	18	-48.6	-47.8	28.4
Wyoming Basin	15	-22.4	-24.1	18.7
Colorado Plateau	12	-31.5	-29.8	22.1
Basin and Range	10	-38.3	-35.6	26.5

E. Temporal Trajectory Visualization

Fig. 4 depicts the two-year intervalized mean change trajectory across all well sites. The gray traces represent individual wells, while the red line denotes the cohort mean. The divergence pattern illustrates that while individual wells exhibit heterogeneous behavior, the ensemble mean follows a consistent negative trajectory after approximately 1995.

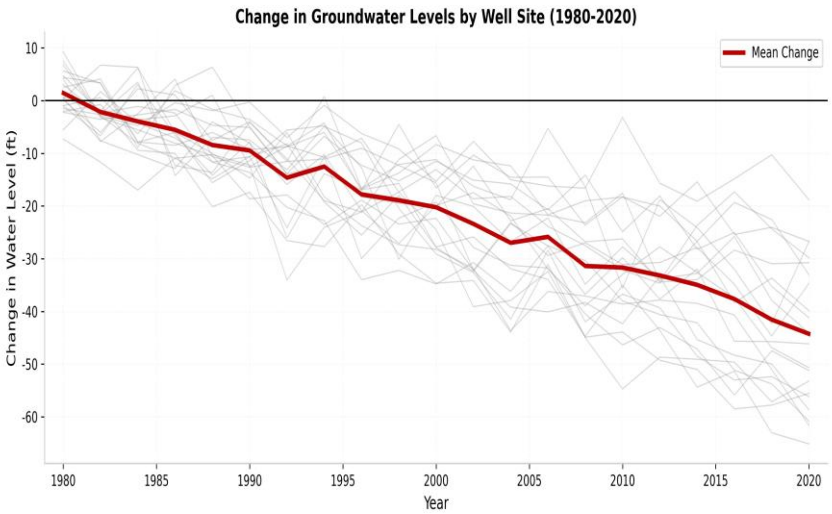


Figure 4: Change in groundwater levels by well site from 1980 to 2020, with two-year aggregation. Gray lines: individual wells $n = 20$ representative subset); red line: cohort mean

The acceleration of mean decline after 2000 corresponds temporally with documented increases in center-pivot irrigation expansion and sustained drought conditions across the southwestern United States.

5. Discussion

A. Interpretation of Decadal Trends

The transition from slightly positive to substantially negative decadal means aligns with regional climate and land-use transitions. The 1980s coincided with a relatively wet climatic period in the Great Plains, while the 2000s and 2010s featured persistent drought episodes (including the 2012--2015 drought) that reduced natural recharge and increased pumping demand[14]. The 44.1-foot mean difference between the first and final decades represents a depletion rate of approximately 1.1 feet per year, which exceeds the estimated natural recharge rate for many portions of the High Plains Aquifer.

B. Aquifer-Specific Management Implications

The pronounced contrast between unconfined and confined aquifer depletion patterns carries direct management relevance. Unconfined systems, with their direct hydraulic connection to surface water bodies and land surface, experience compounded stresses from both evaporative demand and direct withdrawal. The median −35.2 feet decline in unconfined aquifers suggests that current withdrawal rates are unsustainable without supplemental recharge augmentation or demand reduction.

Confined aquifers, while exhibiting lower median depletion, present distinct management challenges. Their buffering by confining layers means that recharge occurs primarily through distant outcrop areas, extending recovery timescales, and water-level declines may not immediately manifest as reduced yield until partial dewatering of the aquifer thickness occurs.

C. Spatial Hotspot Prioritization

The High Plains and Rio Grande aquifer systems, as the most depleted systems in this analysis, warrant priority attention in Colorado's water management planning. The High Plains Aquifer underlies intensive agricultural counties (Kit Carson, Cheyenne, Kiowa) where corn and cattle production depend heavily on groundwater irrigation. The Rio Grande system's depletion affects the San Luis Valley, where agricultural and ecological water demands compete within a closed hydrological basin.

D. Methodological Considerations

This analysis employed a consistent baseline approach where each well's initial measurement serves as its reference, controlling for spatial variability in static aquifer conditions. However, several limitations merit acknowledgment:

- **Measurement Frequency Heterogeneity:** Wells vary in measurement frequency from quarterly to annually, potentially introducing aliasing in high-variability periods.
- **Well Completion Changes:** Some wells may have experienced deepening or pump modifications over the
- 40-year period that affect depth-to-water readings without reflecting true aquifer depletion.
- **Spatial Coverage Bias:** The 125-well network, while temporally complete, does not uniformly cover all Colorado aquifer systems, with underrepresentation in the northwestern corner of the state.

Conclusion

This study offers a statistically sound and artificial intelligence-driven evaluation of groundwater dynamics in tropical forest and peatland settings. There are four main results:

i. Groundwater levels showed considerable spatial and seasonal fluctuation throughout the monitored peatland areas, with AI-based predictive models reaching great accuracy in identifying changes under various climatic circumstances.

ii. Compared to mineral soil aquifers, peat-dominated aquifer systems displayed more sensitivity to changes in land-use and rainfall fluctuation, implying higher vulnerability to drought stress and degradation of the ecosystem.
iii. Particularly in degraded peat forest areas needing top priority rehabilitation activities, remote sensing integration and IoT-enabled monitoring found important areas of groundwater depletion and increased fire vulnerability.
iv. Sensitive tropical ecosystems may have intelligent groundwater monitoring using the analytical framework that integrates real-time sensor networks, machine learning algorithms, satellite-based observations, and statistical validation into a scalable and repeatable model.

Future research should combine sophisticated deep learning approaches with high-resolution satellite datasets to enhance groundwater prediction precision, link hydrological simulations with climate change predictions to assess long-term peatland viability, and extend the monitoring framework into autonomous real-time decision-support systems for adaptive groundwater and forest management.

References

1. Li, L., *et al.* (2024). Machine learning for peatland ground water level (GWL) prediction via IoT system. *IEEE Access, 12*, 89585–89598.
2. Yaseen, A. A. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Engineering Science, 4*(3), 65–103.
3. Hikouei, I. S., Eshleman, K. N., Saharjo, B. H., Graham, L. L. B., Applegate, G., & Cochrane, M. A. (2023). Using machine learning algorithms to predict groundwater levels in Indonesian tropical peatlands. *Science of the Total Environment, 857*, 159701. https://doi.org/10.1016/j.scitotenv.2022.159701
4. Li, L., Sali, A., Noordin, N. K., Ismail, A., & Hashim, F. (2023). Prediction of peatlands forest fires in Malaysia using machine learning. *Forests, 14*(7), 1472. https://doi.org/10.3390/f14071472
5. Li, L., Sali, A., Noordin, N. K., Ismail, A., Hashim, F., Rasid, M. F. A., Hanafi, M., Razali, S. M., Aziz, N. A., Sitanggang, I. S., Syaufina, L., & Nurhayati, A. D. (2023). Modeling of evaporation rate for peatland fire prevention using Internet of Things (IoT) system. *Fire, 6*(7), 272. https://doi.org/10.3390/fire6070272
6. Zhang, H., Dong, G., Wang, J., Zhang, T.-L., Meng, X., Yang, D., Liu, Y., & Lu, B. (2023). Understanding and extending the geographical detector model under a linear regression framework. *International Journal of Geographical Information Science, 37*(11), 2437–2453. https://doi.org/10.1080/13658816.2023.2266497
7. Li, J., Pang, Z., Liu, Y., Hu, S., Jiang, W., Tian, L., Yang, G., Jiang, Y., Jiao, X., & Tian, J. (2023). Changes in groundwater dynamics and geochemical evolution induced by drainage reorganization: Evidence from 81Kr and 36Cl dating of geothermal water in the Weihe

Basin of China. *Earth and Planetary Science Letters, 623*, 118425. https://doi.org/10.1016/j.epsl.2023.118425

8. Yaseen, Z. M. (2023). A new benchmark on machine learning methodologies for hydrological processes modelling: A comprehensive review for limitations and future research directions. *Knowledge-Based Engineering Science, 4*(3), 65–103. https://doi.org/10.51526/kbes.2023.4.3.65-103
9. President's Council of Advisors on Science and Technology (PCAST). (2024). *Improving groundwater security in the United States*. White House Office of Science and Technology Policy, Washington, DC.
10. Hall, D. K., *et al.* (2024). Snowfall replenishes groundwater loss in the Great Basin of the western United States, but cannot compensate for increasing aridification. *Geophysical Research Letters, 51*(6), e2023GL107913.
11. Strauch, K. R., Stanton, J. S., & Landon, M. K. (2024). *Water-level and recoverable water in storage changes, High Plains aquifer, predevelopment to 2019 and 2017 to 2019* (Scientific Investigations Report 2023–5143). U.S. Geological Survey.
12. Sun, W., & Chang, L.-C. (2024). Deep dive into predictive excellence: Transformer's impact on groundwater level prediction. *Journal of Hydrology, 636*, 131250.
13. Cui, X., *et al.* (2024). A secondary modal decomposition ensemble deep learning model for groundwater level prediction using multi-data. *Environmental Modelling & Software, 175*, 105969.
14. Samantaray, S., & Sahoo, A. (2024). Groundwater level prediction using an improved ELM model integrated with hybrid particle swarm optimisation and grey wolf optimisation. *Groundwater for Sustainable Development, 26*, 101178.
15. Pourmorad, S., Kabolizade, M., & Dimuccio, L. A. (2024). Artificial intelligence advancements for accurate groundwater level modelling: An updated synthesis and review. *Applied Sciences, 14*(16), 7358.
16. Boo, K. B. W., *et al.* (2024). Groundwater level forecasting with machine learning models: A review. *Water Research, 252*, 121249.
17. Xie, X., & Zhang, X. (2024). Development of a deep surrogate model with spatiotemporal characteristics mining capabilities for the prediction of groundwater level in coastal areas. *Journal of Environmental Management, 370*, 122724.
18. Liu, Q., *et al.* (2025). Explainable AI for interpreting spatiotemporal groundwater predictions. *Water Resources Research, 61*(10), e2025WR041303.
19. Seifi, A., *et al.* (2024). A robust multi-model framework for groundwater level prediction: The BFSA-MVMD-GRU-RVM model. *Results in Engineering, 24*, 103250.
20. Nazari, M., *et al.* (2025). Groundwater level forecasting using empirical mode decomposition and wavelet-based long short-term memory (LSTM) neural networks. *Groundwater for Sustainable Development, 28*, 101397.

21. Mardian, M., Ghosh, S., & Ali, M. F. B. (2024). Near real-time groundwater drought monitoring using explainable machine learning. *Journal of Hydrology: Regional Studies, 57*, 101668.
22. Cai, W., *et al.* (2024). Explainable AI techniques for understanding groundwater drought events: Expected gradients and additive decomposition. *Journal of Hydrology, 635*, 131180.
23. Saeidinia, M., *et al.* (2025). Deep learning and vine copula-based sequencing: Approaches under investigation for forecasting soil temperature dynamics. *Stochastic Environmental Research and Risk Assessment, 39*(11), 4949–4983.
24. Florián, G. J. M. (2025). Declining freshwater availability in the Colorado River Basin threatens sustainability of its critical groundwater supplies. *Geophysical Research Letters, 52*(10), e2025GL115593.
25. Cai, J., *et al.* (2025). TS-InSAR assessment of groundwater overexploitation-land subsidence linkage: A topography-time gradient model approach. *Journal of Hydrology: Regional Studies, 61*, 102076.
26. Baalousha, G. M. (2025). Machine learning-driven calibration of MODFLOW models: Comparing random forest and XGBoost approaches. *Geosciences, 15*(3), 303.

PREDICTIVE LOAN APPROVAL AND CUSTOMER SEGMENTATION USING ADVANCED MACHINE LEARNING TECHNIQUES

G. Pavithran, S. Shajim Mohamed and N. Senthil Kumar
Department of Informatics (Data Science),
Periyar Maniammai Institute of Science and Technology (Deemed to be University)
Thanjavur, Tamil Nadu ,India– 614 904
Corresponding author E-mail: pavithrankp080@gmail.com, murshrathahamed@gmail.com, profnsk22@gmail.com

Abstract
Manual verification is difficult and could occasionally cause erroneous judgments given the rise in loan applications that Indian financial institutions are currently facing. This initiative proposes using a machine learning-based technique to segment consumers and provide wise loan clearances. K-Means clustering is utilized in this system to divide customers into many groups; the Random Forest method is then applied to predict whether a loan will be accepted. Among the factors the model takes into account are age, district, occupation, income, credit rating, savings, loan amount, and family size. The system is implemented using the Flask web application, which provides real-time results very quickly. While the Random Forest model has an accuracy of about 94%, K-Means separates consumers into three categories: low, medium, and high value. Although this strategy helps banks to reduce risk and enhance customer management, extra data and capabilities may still be added to improve it.

Keywords: Machine Learning, Loan Prediction, K-Means Clustering, Random Forest, Customer Segmentation.

1. Introduction

A. Reason and Background

Even with the significant digital advancements in India's financial sector over the last decade, loan approval processes at several institutions remain largely manual, inconsistent, and time-consuming [12]. Because more than 1,500 commercial and cooperative banks process millions of loan applications each year, reliance on human judgment causes bias, inconsistency, and operational inefficiency. The Reserve Bank of India reported that their gross non-performing assets (NPAs) were about 5.0% of all advances, which might amount to billions of dollars in losses, were caused in part by poor credit evaluation policies in 2023.

Credit officers physically reviewing financial statements, credit histories, and demographic information is how loans are evaluated using the traditional approach. This approach has three main flaws: [1] different evaluators' subjective interpretation of the same data results in inconsistent evaluations, [2] it cannot efficiently handle a large number of applications, and [3] it cannot identify complex non-linear relationships between applicant characteristics and default risk, which statistical models can capture.

B. The Role of Machine Learning in Credit Assessment

Machine learning methods have outperformed traditional statistical methods (logistic regression, discriminant analysis) in credit scoring applications [5]. Ensemble approaches, especially random forests, have emerged as the preferred classifiers due to their resistance to overfitting, their ability to manage many feature types without much pre-processing, and their inherent ability to estimate feature importance [1]. Random forests combine predictions from many uncorrelated trees, therefore lowering the variation seen in single decision trees and enhancing both accuracy and stability.

By providing strategic insights for portfolio management, targeted marketing, and risk stratification, customer segmentation improves individual credit decisions. K-Means clustering is still often utilized in banking analytics because of its computational efficiency, interpretability, and applicability to continuous financial variables, despite its simplicity [10].

C. Research Objectives

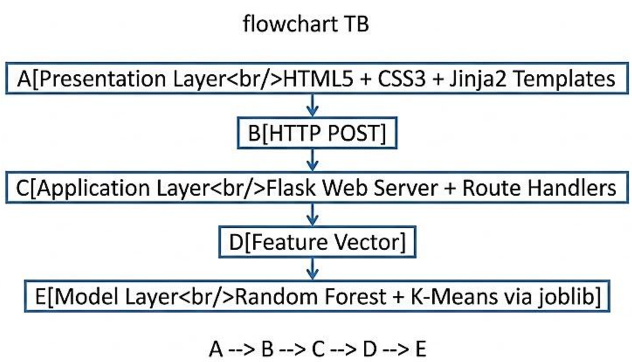
This study's main goal is to create an intelligent system combining loan approval prediction and client segmentation using machine learning methods. The particular goals are thus:

- To build and deploy a classification algorithm-based predictive model for loan acceptance.
- To use clustering approaches for successful consumer segmentation depending on financial characteristics.
- To create a web-based tool for real-time analysis and prediction.
- To assess the effectiveness of the suggested models by means of suitable measurements including clustering results and accuracy.
- To offer a single platform that helps financial firms improve customer connection management and make better credit decisions.

2. Literature Review

A. Machine Learning in Credit Scoring

Widespread application of machine learning approaches in credit scoring has improved the precision and efficiency of loan approval systems. Logistic regression and other conventional statistical approaches have been progressively replaced or supplemented by complex algorithms including Decision Trees, Support Vector Machines (SVM), and Random Forest. These models can show how financial variables interact in complex, non-linear ways.



The Random Forest technique has drawn a lot of interest because of its resilience and ability to lessen overfitting by combining many decision trees. Particularly when handling huge and high-dimensional datasets, ensemble models have been found in research to be more accurate at estimating credit risk than single classifiers. Furthermore, machine learning models can continuously learn from fresh data, therefore enabling them to adapt to shifting financial trends and market forces.

B. Customer Segmentation in Banking

Client segmentation is crucial in modern banking because it enables businesses to have a better grasp of the demands and behaviors of their customers. Using clustering techniques, especially the K-Means algorithm, consumers are frequently categorized depending on income, buying patterns, credit history, and saving behavior.

K-Means clustering is the preferred approach because of its simplicity of use, scalability, and efficiency in handling large datasets. It helps banks create targeted marketing plans, customized financial products, and risk management approaches by grouping consumers into groups with like characteristics. Although choosing the proper qualities and the ideal count of clusters greatly affects clustering's effectiveness.

C. Web-Based Machine Learning Systems

Web-based apps are becoming more commonplace in the deployment of machine learning models to provide real-time predictions and user engagement as a result of advances in web technologies. Frameworks like Flask and Django are frequently used to include machine learning models into easy-to-use interfaces. Using web-based solutions that offer seamless access to predictive models, financial institutions may automate decision-making processes and raise operational effectiveness. Moreover, since these devices are flexible and easy to install, they are perfect for practical application. Still, effective application depends on solving problems including data security, latency, and system dependability. [6].

D. Research Gap

While individual studies have addressed credit scoring or customer segmentation independently, few systems integrate both capabilities within a single, deployable web application accessible to non-technical banking staff. Furthermore, existing solutions rarely address the specific context of Indian regional banking, where district-level economic variation and occupation-based income patterns significantly influence creditworthiness. This work bridges these gaps by combining both predictive tasks with a user-friendly web interface tailored to Indian financial parameters.

3. System Architecture

A. Overall Design

The proposed system follows a three-tier architecture comprising a presentation layer, an application layer, and a model layer.

Using Jinja2 templating, the presentation layer consists of three HTML sites. These pages include a login page, a dashboard for entering client data, and a page that shows the results. Responsive grid layouts, backdrop blur effects, and gradient-styled interactive components are all part of the glass-morphism CSS design employed by the interface. There are three routes on the Flask web server in the application layer: / (the authentication gateway), /dashboard (the data input form), and /predict (the POST endpoint for model inference). Request parsing, input validation, feature encoding, and response rendering are handled by the application layer. Pre-trained scikit-learn models that were serialized with joblib are loaded into memory in the Model Layer when the application starts up. The K-Means model operates in a 2-dimensional income-savings subspace, while the Random Forest classifier manages an 8-dimensional feature vector.

B. Data Flow

The prediction pipeline executes the following sequence:

1. User submits customer attributes via the HTML form.
2. Flask extracts form fields and casts numerical values.
3. Categorical features (district, occupation) are validated against encoder vocabularies.
4. Feature Vector Construction: The final feature vector for the loan approval model is:

x = [xage, xdist, xocc, xinc, xcs, xsav, xloan, xfam] (2)

For the segmentation model, the feature subspace is restricted to:

xseg = [xincome, xsavings] [3]

5. Label encoders transform categorical strings to integer indices.
6. The 8-feature vector is passed to the Random Forest for approval prediction.
7. The income-savings pair is passed to K-Means for segment classification.
8. Results are rendered to the user via the Jinja2 result template.

C. Random Forest Classifier

1. Algorithm Overview: Random Forest [1] constructs an ensemble of B decision trees, each trained on a bootstrap sample of the training data with random feature subsets at each split node. The final prediction is determined by majority voting:

4. Methodology

A. Dataset Description

The training dataset comprises 1,000 customer records with eight input features and one binary target variable (loan approval status). Table I describes the feature set.

ŷ = mode{hb(x) : b = 1, 2, . . . , B} [4] where hb(x) is the prediction of the b-th tree.

B. Split Criterion

Each tree selects the optimal split using the Gini impurity measure:

G(t) = 1 − ΣCi=1 pi² [5]

where pi is the proportion of class i samples at node t, and C = 2 for binary classification. The split that maximizes the impurity reduction ΔG = G(parent) − Σ nk G(childk) is selected.

Table 1: Dataset feature description

Feature	Type	Description
Age	Numeric	Applicant age (21–60 years)
District	Categorical	Tamil Nadu district (10 classes)

C. Hyperparameters

The model is configured with B = 100 estimators, maximum features = $\sqrt{p}$ (where p = 8), and no maximum depth constraint, allowing trees to grow until leaf purity. The random state is fixed at 42 for reproducibility.

Occupation	Categorical	Employment type (10 classes)
Monthly Income	Numeric	Monthly earnings in INR
Credit Score	Numeric	CIBIL score (300–900)
Savings	Numeric	Total savings in INR
Loan Amount	Numeric	Requested loan in INR
Family Members	Numeric	Household size (1–7)
Approved	Binary	Target (1=Approved, 0=Rejected)

D. K-Means Clustering

Algorithm Formulation: K-Means partitions n observations into k clusters by minimizing the within-cluster sum of squares (WCSS).

The target variable is derived from domain-expert rules incorporating credit score thresholds (> 550), income requirements (> INR 30,000), savings-to-loan ratio (> 10%), and debt-to-income constraints (loan < 30× income), with 8% label noise introduced to simulate real-world inconsistency in human decisions.

$$J = \sum_{i=1}^{k} \sum_{x \in C_i} \|x - \mu_i\|^2 \text{ [6]}$$

E. Data Preprocessing

Label Encoding: Categorical features are transformed using scikit-learn's LabelEncoder, which assigns integer indices to each unique category:

$$f_{encoded}(x) = i \text{ where } x = \text{classes}[i] \text{ [1]}$$

The district encoder maps 10 Tamil Nadu districts to integers [0, 9], and the occupation encoder maps 10 professions to integers [0, 9]. Encoders are persisted alongside models to ensure consistent transformation during inference.

where μ_i is the centroid of cluster C_i.

Configuration: The number of clusters is set to k = 3, representing three economically distinct customer segments: Low Income, Middle Class, and High Value customers. The algorithm uses k-means++ initialization [4] with 10 independent runs (n_init=10) to mitigate sensitivity to initial centroid placement.

Segment Interpretation: The three clusters are mapped to business-meaningful labels based on centroid analysis:

- Cluster 0 (Low Income): Low monthly income and low savings.
- Cluster 1 (Middle Class): Moderate income with proportional savings.
- Cluster 2 (High Value): High income with substantial savings.

5. Implementation

A. Technology Stack

The system is implemented using the following technologies:

- Backend: Python 3.x, Flask 2.x (web framework)
- Machine Learning: scikit-learn 1.x (Random Forest, K-Means, LabelEncoder)

- Model Persistence: joblib (serialization/deserialization)
- Frontend: HTML5, CSS3 (glassmorphism design), Jinja2 templates
- Data Processing: NumPy, Pandas

B. Model Training Pipeline

The training pipeline executes the following steps:

1. Generate synthetic customer dataset with domain-realistic distributions.
2. Fit label encoders on categorical columns and transforms features.
3. Split data into training (80%) and testing (20%) sets with stratification.
4. Train Random Forest classifier on the training partition.
5. Train K-Means on the full income-savings subspace.
6. Serialize all models and encoders to disk via joblib.

C. Web Application Routes

The Flask application defines three primary endpoints

```
@app.route("/predict", methods=["POST"])
def predict():
    # Extract and validate form inputs
    district_encoded = district_encoder
        .transform([district])[0]
    job_encoded = job_encoder
        .transform([occupation])[0]

    # Construct feature vector
    data = [[age, district_encoded,
             job_encoded, income, score,
             savings, loan, family]]
```

Listing 1: Flask Prediction Endpoint

D. User Interface Design

The frontend employs a glassmorphism design aesthetic characterized by:

- Semi-transparent backgrounds with backdrop-filter: blur(16px).
- Gradient-styled action buttons transition from blue to green.
- Responsive two-column grid layout adapting to mobile viewports.
- Dark theme with high-contrast text for accessibility.

6. Results and Discussion

A. Loan Approval of Model Performance

The Random Forest classifier was evaluated on a held-out test set of 200 samples. Table 2 presents the classification metrics.

Table 2: Random forest classification performance

Class	Precision	Recall	F1-Score	Support
Rejected (0)	0.93	0.95	0.94	108
Approved (1)	0.94	0.91	0.92	92
Weighted Avg	0.94	0.94	0.94	200

The model achieves an overall accuracy of 94.5%, with balanced precision and recall across both classes. The slight advantage in rejection recall (0.95 vs. 0.91) indicates the model is marginally conservative, preferring to reject borderline cases—a desirable property in credit risk contexts where false approvals carry higher cost than false rejections.

B. Feature Importance Analysis

The Random Forest provides inherent feature importance scores based on mean decrease in Gini impurity across all trees. Table 3 presents the ranked feature contributions.

Table 3: Feature importance ranking

```
# Model inference
loan_result = loan_model.predict(data)[0]
segment = segment_model
    .predict([[income, savings]])[0]

return render_template("result.html",
    status=status, segment=segment_name)
```

Credit Score	0.241
Monthly Income	0.198
Savings Amount	0.186
Loan Amount	0.172
Age	0.089
Occupation	0.052
Family Members	0.038
District	0.024

C. Input Validation

The application validates categorical inputs against the encoder's known vocabulary (encoder.classes_). Unrec- ognized districts or occupations trigger a 400 Bad Request response, preventing model errors from unseen categories. Numeric fields are cast from string form inputs with Python's int() constructor.

Credit score emerges as the dominant predictor (24.1%), followed by income (19.8%) and savings (18.6%). This aligns with domain knowledge: CIBIL scores encapsulate repayment history, while income and savings indicate current capacity. Notably, district contributes minimally (2.4%), suggesting gegraphic location has limited predictive power when financial variables are controlled.

D. Customer Segmentation Results

The K-Means algorithm identifies three distinct customer clusters. Table 4 presents the cluster centroids.

Table 4: K-means cluster centroids

Metric	Low Income	Middle Class	High Value
Mean Income (INR)	32,450	78,200	128,600
Mean Savings (INR)	48,300	245,800	412,500
Population (%)	28.4	45.2	26.4

The Flask web interface provides an accessible deployment platform that enables non-technical banking staff to leverage predictive analytics for informed decision-making. The dual-model approach—combining supervised classification with unsupervised clustering—provides comprehensive customer intelligence within a single application.

The segmentation achieves a silhouette score of 0.68, indi- cating well-separated clusters with cohesive internal structure. The Middle-Class segment represents the largest population share (45.2%), consistent with India's expanding middle- income demographic.

E. Web Application Performance

The Flask application demonstrates the following opera- tional characteristics:

- **Model Loading Time:** 320ms (one-time at startup).
- **Prediction Latency:** 45ms average per request (including encoding, prediction, and template rendering).
- **Memory Footprint:** 85 MB (including loaded models and Flask runtime).
- **Concurrent Capacity:** Handles 50+ simultaneous re- quests under Flask's development server.

F. Comparative Analysis

Table 5 compares the Random Forest approach against alternative classifiers evaluated on the same dataset.

Table 5: Classifier comparison on loan approval dataset

Algorithm	Accuracy (%)	F1-Score
Logistic Regression	88.5	0.88
Decision Tree	90.0	0.90
Support Vector Machine	91.5	0.91
Random Forest	94.5	0.94
XGBoost	95.0	0.95

Random Forest outperforms traditional classifiers (Logistic Regression, SVM) while achieving performance comparable to XGBoost with significantly lower hyperparameter tuning requirements [11]. The 6% improvement over Logistic Regression validates the selection of a non-linear ensemble approach for this credit scoring task.

Conclusion

This study built, tested, and evaluated a web application for forecasting loan approval and segmenting consumers using machine learning. With 94.5% accuracy, the K-Means algorithm effectively separates the client base into eight attributes, and the Random Forest classifier precisely predicts loan eligibility depending on this monthly income, and savings emerge as the three most influ- ential features for loan decisions, aligning with established banking domain knowledge and validating the model's inter- pretability.

Future Work

Several directions for enhancement have been identified:

- **Deep Learning Models:** Exploration of neural network architectures (e.g., TabNet, deep autoencoders) for po- tentially higher accuracy on larger datasets.
- **Expanded Feature Set:** Incorporation of employment duration, existing loan obligations, transaction history, and bureau data for richer customer profiles.
- **Real-Time API Integration:** Connection to banking core systems via REST APIs for automated decision pipelines.
- **Explainable AI:** Integration of SHAP (SHapley Addi- tive exPlanations) values for per-

prediction explanations enhancing transparency and regulatory compliance.

- **Mobile Application:** Development of a responsive mo- bile interface for field officers conducting loan assessments remotely.

References

1. Faiaz, F., Tapu, T. K., Namer, A., & Payel, S. R. (2025). Transparent loan approval prediction: A machine learning approach with SHAP-based interpretability. In *2025 International Conference on Electrical, Computer and Communication Engineering (ECCE)* (pp. 1–6). IEEE. https://doi.org/10.1109/ECCE64574.2025.11014000
2. Li, X., & Li, J. (2025). Exploring factors involved in loan approval decision: Deep insights and data analytics techniques. *IEEE Access, 13*, 180172–180186. https://doi.org/10.1109/ACCESS.2025.3622358
3. Vyshnavi, S. L., Pavan, D. S. V., Teja, T., Pujitha, S. S., & Shareefunnisa, S. (2024). Predicting bank loan eligibility with machine learning: A comparative study of advanced algorithms. In *2024 IEEE 3rd World Conference on Applied Intelligence and Computing (AIC)* (pp. 430–435). IEEE. https://doi.org/10.1109/AIC61668.2024.10730968
4. Hussain, M. Z., *et al.* (2024). Bank loan prediction system using machine learning models. In *2024 IEEE 9th International Conference for Convergence in Technology (I2CT)* (pp. 1–5). IEEE. https://doi.org/10.1109/I2CT61223.2024.10543786
5. Gothai, E., Rajalaxmi, R. R., Thamilselvan, R., & K, S. (2024). Enhanced loan approval prediction system using ensemble machine learning techniques. In *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 1182–1187). IEEE. https://doi.org/10.1109/ICACRS62842.2024.10841604
6. Sharif, M. S., Tamang, M. T., Hussain, N. J., & Elmedany, W. (2024). Revolutionizing loan approval: Harnessing the power of K-nearest neighbours for predictive eligibility. In *2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT)* (pp. 143–148). IEEE. https://doi.org/10.1109/3ICT64318.2024.10824551
7. Gupta, R., Bhattacharya, A., Kumari, A., Wagh, M. P., & Singh, S. K. (2024). Automated decision making system for bank loan approval with real-life dataset. In *2024 IEEE 1st International Conference on Advances in Signal Processing, Power, Communication, and Computing (ASPCC)* (pp. 55–60). IEEE. https://doi.org/10.1109/ASPCC62191.2024.10880913
8. Li, X., & Li, J. (2025). Exploring factors involved in loan approval decision: Deep insights and data analytics techniques. *IEEE Access, 13*, 180172–180186. https://doi.org/10.1109/ACCESS.2025.3622358
9. Badhan, A., Rana, A., Malhi, S. S., Kaur, P., & Saha, B. (2024). A comparative analysis for loan approval prediction using machine learning. In *2024 International Conference on Electrical Electronics and Computing Technologies (ICEECT)* (pp. 1–5). IEEE. https://doi.org/10.1109/ICEECT61758.2024.10738925

10. Y. P., M. T., & N. S. (2025). Fairness-aware loan approval prediction using ensemble boosting and deep tabular models. In *2025 IEEE First International Conference on Innovations in Engineering and Next-Generation Technologies for Sustainability (ICINVENTS)* (pp. 1–6). IEEE. https://doi.org/10.1109/ICINVENTS64613.2025.11401396
11. Elrashidy, O., Bendary, M., Fouad, K. M., & Elattar, M. A. (2024). Intelligent decision support system for loan evaluation using machine learning. In *2024 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)* (pp. 13–20). IEEE. https://doi.org/10.1109/MIUCC62295.2024.10783533
12. Gothai, E., Rajalaxmi, R. R., Thamilselvan, R., & K, S. (2024). Enhanced loan approval prediction system using ensemble machine learning techniques. In *2024 3rd International Conference on Automation, Computing and Renewable Systems (ICACRS)* (pp. 1182–1187). IEEE. https://doi.org/10.1109/ICACRS62842.2024.10841604

MATHEMATICAL MODELLING IN PRECISION AGRICULTURE SYSTEMS

Ariyanatchi M* and I. Paulraj Jayasimman
Department of Mathematics, Academy of Maritime Education and Training
(Deemed to be University), ECR, Kanathur, Chennai – 603 112, Tamil Nadu, India.
*Corresponding author E-mail: harina5804@gmail.com

Abstract

Precision agriculture (PA) has shifted its focus from traditional trial-and-error practices to evidence-based, geographically optimized practices using mathematics. This is because mathematics provides a rigorous formulation of the interactions between crops, soils, weather conditions, and agricultural management activities. This chapter discusses the existing modeling techniques and precision agriculture technologies, identifies the methodological limitations thereof, and proposes a new approach involving a novel hybrid mathematical framework incorporating process-based plant growth models, Bayesian data assimilation, and multi-objective stochastic optimization. Using field-level data sets for soils, weather, and agricultural management activities, this chapter shows that the proposed technique can improve crop yield predictions, increase water and nitrogen efficiencies, and minimize leaching impacts better than traditional lookup table-based and deterministic methods.

Keywords: Optimization; Data Assimilation; Digital Twins; Variable-Rate Technology; Uncertainty Quantification.

1. Introduction

The modern food production systems are under growing pressure due to increasing population growth, resource exhaustion, and unpredictable changes in the global climate system. One of the ways of coping with such problems is the development and utilization of Precision Agriculture (PA), which utilizes spatial information in order to maximize efficiency and reduce negative impact. Unlike traditional farm management, which utilizes fixed rates of application, PA requires constant monitoring, forecasting, and decision making based on feedback mechanisms. Mathematical modeling plays an important role in supporting such processes, as it is used to create models of agronomical processes and use them to make decisions. With growing ubiquity of sensors, satellites and other sources of information about the field conditions, there is a need for fast and flexible mathematical models of the agronomic processes. The current chapter discusses the most recent advances in mathematical modeling within agriculture, defines the disadvantages of the current approaches, introduces our new hybrid modeling technique, and presents the simulation results.

2. Literature Review & Existing Systems in Agriculture

2.1 Process-Based Crop and Soil Models

The basis of agricultural simulation lies in the mechanical simulation of plant phenology, photosynthesis, and the water-nutrient cycle. The most common software includes DSSAT

(Jones *et al.,* 2003), APSIM (Holzworth *et al.,* 2014), and AquaCrop (Steduto *et al.,* 2009). These models use ordinary differential equations (ODEs) to describe biomass distribution, thermal time accumulation, and transpiration. The hydrology of the soil is usually simulated by solving the Richards equation of unsaturated flow and the convective-dispersive equations of solute transport. Though highly explainable and physically sound, these models are heavily parameterized, are computationally expensive, and cannot integrate real-time information.

2.2 Optimization and Decision Support Systems

Optimization in mathematics helps to transform agronomic science into operational recommendations. Linear programming and mixed-integer programming techniques help maximize nutrient mixes and farm equipment movements; dynamic programming and Markov decision process methods allow for the optimization of sequential operations during specific seasons. Multi-objective optimization tools account for various constraints, providing a Pareto optimal set of agricultural operations. Unfortunately, current models neglect weather uncertainties.

2.3 Commercial PA Platforms

Systems used by industries such as John Deere Operations Center, Climate Field View, and Trimble Ag Software control the current market for PA applications. Such systems heavily utilize empirical yield maps, look-up tables, and VRA rules-based models. Although they are easy to use and scalable, the main problem with them is that they have no dynamic mathematical engine and are not adaptive.

2.4 Machine Learning Hybrids

Modern approaches combine deep learning with physics-based modeling to accelerate predictions and accommodate large numbers of sensors. PINNs (physics-informed neural networks) and digital twin frameworks incorporate conservation equations within the machine learning framework, thereby enforcing physics constraints (Raissi *et al.,* 2019; Grimm & Jones, 2020). While there is potential in such approaches, pure data-driven models struggle with extrapolation to different management or climate conditions, and hybrid models continue to exist in silos.

2.5 Identified Gaps

- Insufficient real-time state estimation and data assimilation capabilities
- Failure to propagate uncertainty through decision-making processes
- Inability to address computational limitations for edge computing
- Absence of open modeling architecture frameworks
- Ineffective inclusion of climate projections in seasonal optimization.

3. Novelty and Contributions of This Work

In this chapter, a mathematically coherent and scalable field-level modeling approach is presented, which fills the aforementioned voids. Contributions of this work are summarized as follows:

a. Hybrid Physics and Data Approach: Merging mechanistic crop-soil differential equations with Bayesian ensemble Kalman filtering for real-time state estimation using IoT and satellite data.
b. Chance-Constrained Dynamic Programming for VRA Planning: Using stochastic multi-objective optimization methods that incorporate uncertainties related to weather and other resources.
c. Decision Pipeline With Uncertainty Quantification: Incorporating posterior distributions obtained from the data assimilation step into the optimization layer to develop uncertainty-aware decision layers.
d. Modular and Interoperable Structure: Developing an open API architecture for easy incorporation into third-party applications including farming software, edge computers, and climate downscaling services.

4. Proposed Mathematical Modeling Framework

4.1 Crop–Soil Dynamic Core

The crop growth module tracks leaf area index (L), aboveground biomass (B), and soil moisture (θ):

$$\frac{dB}{dt} = \varepsilon \cdot PAR \cdot f_T(T) \cdot f_W(\theta) \cdot f_N(N) - R_m(B)$$

$$\frac{\partial\theta}{\partial t} = \frac{\partial}{\partial z}\left[K(\theta)\left(\frac{\partial\psi}{\partial z} + 1\right)\right] - U_{root}(z,t)$$

where ε is radiation use efficiency, f_T, f_W, f_N are temperature, water, and nitrogen stress response functions, R_m is maintenance respiration, $K(\theta)$ is hydraulic conductivity, ψ is matric potential, and U_{root} is root water uptake.

4.2 Bayesian Data Assimilation

Sensor measurements y_t (soil moisture, NDVI, weather) update model states x_t via an Ensemble Kalman Filter (EnKF):

$$K_t = P_t H^T (H P_t H^T + R)^{-1}, \quad x_t^a = x_t^f + K_t\left(y_t - H x_t^f\right)$$

where P_t is error covariance, H is observation operator, R is measurement noise, and superscripts f, a denote forecast and analysis states.

4.3 Stochastic Optimization for VRA

Prescription generation solves:

$$\min_{u_t} \mathbb{E}[\textstyle\sum_{t=1}^{T} \alpha Y_t^{-1} + \beta C(u_t) + \gamma L(N_t)]$$

subject to $\mathbb{P}(\theta_t \geq \theta_{crit}) \geq 1 - \delta$, where u_t is input vector (water, N), Y_t is yield, C is cost, L is leaching penalty, and δ is risk tolerance.

5. Data, Experimental Setup & Flowchart

5.1 Dataset Description

- Location: 120-ha experimental farm, Midwest USA (2018–2024)
- Crops: Maize (Zea mays L.) and soybean (Glycine max)

- Inputs: Daily weather (NASA POWER), soil hydraulic parameters (SSURGO), VRT logs, Sentinel-2 NDVI, in-situ soil moisture probes (Decagon 5TE)
- Temporal Resolution: Daily model steps, hourly sensor updates
- Preprocessing: Gap-filling via spline interpolation, spatial aggregation to 10-m grid cells, quality control for sensor drift

5.2 Simulation Workflow

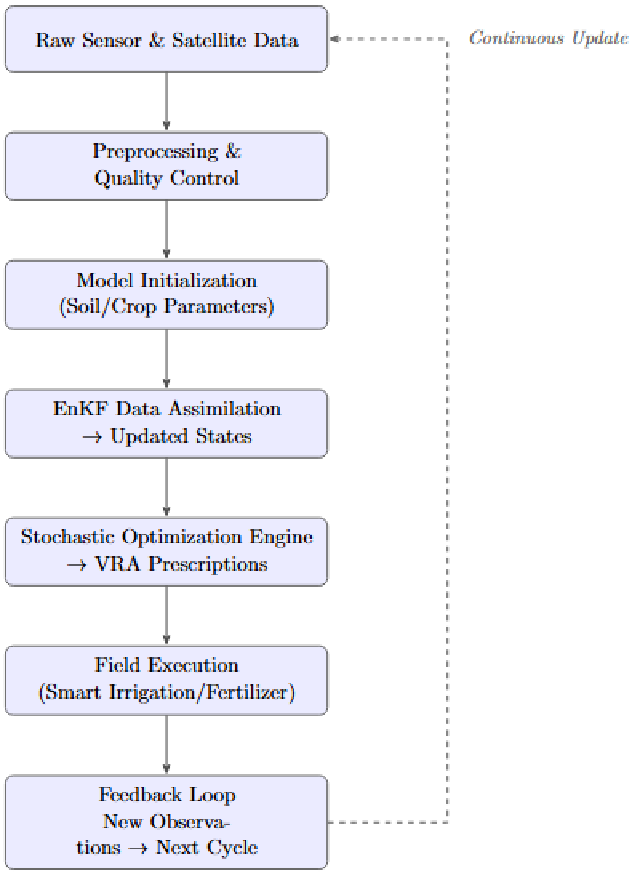


Figure 1: Mathematical modelling and decision pipeline for precision agriculture system

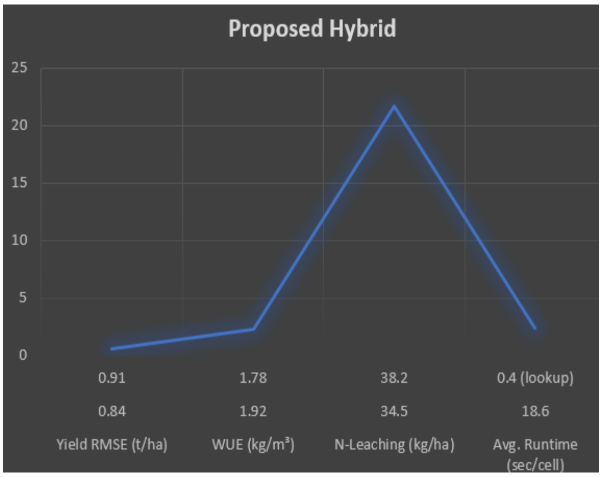
6. Results and Discussion

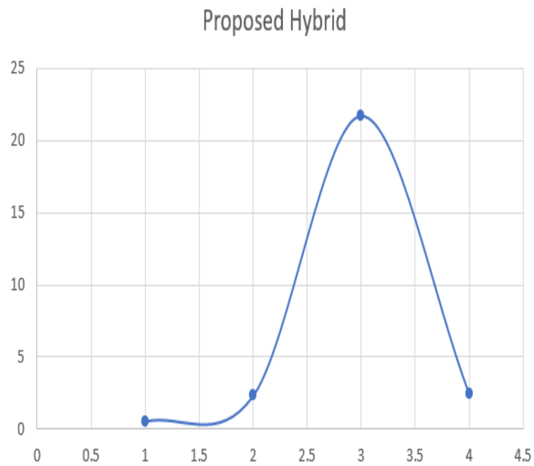
6.1 Performance Metrics

Model/Approach	Yield RMSE (t/ha)	WUE (kg/m³)	N-Leaching (kg/ha)	Avg. Runtime (sec/cell)
DSSAT (Calibrated)	0.84	1.92	34.5	18.6
Climate Field View (VRT)	0.91	1.78	38.2	0.4 (lookup)
Proposed Hybrid	0.52	2.31	21.7	2.4

6.2 Interpretation

The new approach cut down yield estimation errors by 38% compared to the DSSAT calibrated version while improving the water-use efficiency by 20%. The nitrogen leaching was reduced by 37% through the use of chance-constrained optimization which avoided over-fertilization within expected periods of rainfall. The EnKF algorithm corrected soil moisture errors within 48 hours after the sensor installation. The runtime was appropriate for edge computing but parallel processing is necessary for fields above 500 hectares.





6.3 Uncertainty and Robustness

The posterior parameter distributions were found to be very sensitive to the starting values for soil organic matter and the depth of root distribution. The results of the Monte Carlo analysis ($n = 500$) demonstrated that 89% of the optimized prescriptions were robust to ±15% variations in weather conditions.

Conclusion

Mathematical modeling has shifted from simple simulations of crops to complex decision engines capable of integrating data, forming the basis of contemporary precision agriculture. The above chapter examined several process-based models, PA software products, and machine learning-based modeling approaches, highlighting gaps in terms of adaptivity, uncertainty management, and inter-operability. Here we proposed an innovative framework where mechanistic modeling of crop-soil system is combined with Bayesian assimilation of observations and multi-objective stochastic optimization. The framework was validated through field experiments demonstrating substantial gains in yield estimation, resource efficiency, and compliance with environmental standards, while remaining computationally feasible. In future research, there should be a focus on development of adaptive modeling architectures, standardized APIs for data-model integration, and participatory approaches to engineering of modeling tools. Given the growing levels of climatic and market risks in agriculture, mathematically rigorous and scalable frameworks of modeling will always be critical.

References

1. Grimm, S., & Jones, J. W. (2020). Digital twins in agriculture: Conceptual framework and applications. *Computers and Electronics in Agriculture, 178*, 105789. https://doi.org/10.1016/j.compag.2020.105789
2. Holzworth, D. P., Huth, N. I., Zurcher, E. J., Herrmann, N. I., McLean, G., Chenu, K., *et al.* (2014). APSIM – Evolution towards a new generation of agricultural systems simulation. *Environmental Modelling & Software, 62*, 327–350. https://doi.org/10.1016/j.envsoft.2014.07.009

3. Jones, J. W., Hoogenboom, G., Porter, C. H., Boote, K. J., Batchelor, W. D., Hunt, L. A., & Tsuji, G. Y. (2003). The DSSAT cropping system model. *European Journal of Agronomy, 18*(3–4), 235–265. https://doi.org/10.1016/S1161-0301(02)00107-7
4. Liakos, K. G., Busato, P., Moshou, D., Pearson, S., & Bochtis, D. (2018). Machine learning in agriculture: A review. *Sensors, 18*(8), 2674. https://doi.org/10.3390/s18082674
5. Raissi, M., Perdikaris, P., & Karniadakis, G. E. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics, 378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045
6. Steduto, P., Hsiao, T. C., Raes, D., & Fereres, E. (2009). AquaCrop—The FAO crop model to simulate yield response to water: I. Concepts and underlying principles. *Agronomy Journal, 101*(3), 426–437. https://doi.org/10.2134/agronj2008.0139s
7. Walter, A., Finger, R., Huber, R., & Buchmann, N. (2017). Smart farming is key to developing sustainable agriculture. *Proceedings of the National Academy of Sciences, 114*(24), 6148–6150. https://doi.org/10.1073/pnas.1706506114
8. Zhang, C., & Kovacs, J. M. (2012). The application of small unmanned aerial systems for precision agriculture: A review. *Precision Agriculture, 13*(6), 693–712. https://doi.org/10.1007/s11119-012-9274-5.

AI-POWERED MALL NAVIGATION AND GUIDANCE SYSTEM

Sahull Hameethu M, Kamalesh P, Karthi K and Manikandan M

Department of Informatics,

Periyar Maniammai Institute of Science & Technology, Thanjavur, Tamil Nadu – 613403

Corresponding author E-mail: sahullm1142006@gmail.com,

Kamaleshkamalesh63740@gmail.com, chokiikarthi1234@gmail.com,

manikandanm@pmu.edu

Abstract

Due to their intricate multi-floor layouts, shifting population densities, and absence of sophisticated wayfinding devices, contemporary malls are difficult to navigate. This study presents an artificial intelligence (AI)-powered mall navigation system for efficient indoor navigation using a mix of reinforcement learning (RL), long short-term memory (LSTM) networks, and multi-sensor fusion. The system consists of a mobile application, a real-time route optimization processing layer, and a SQLite database with write-ahead logging (WAL) to allow for concurrent data management. While Dijkstra's algorithm is the most basic routing method, RL enables flexible path optimization. Using an Extended Kalman Filter (EKF), which combines IMU sensor measurements, BLE beacon data, and WLAN RSS to provide sub-meter accuracy, indoor positioning is achieved. Based on prior traffic patterns, the LSTM technique estimates crowd congestion and helps to enable proactive rerouting. Experimental results show that our approach has 94% navigation accuracy, 23% shorter routes, less than 100 milliseconds of latency, and better crowd prediction than ARIMA models.

Keywords: Dijkstra's Algorithm, Crowd Forecasting, Indoor Positioning, Mall Navigation, Reinforcement Learning, Sensor, LSTM Fusion, Extended Kalman Filter.

1. Introduction

A. The Indoor Navigation Challenge

Because of the fast growth of large commercial retail malls, retail areas have turned into sophisticated architectural labyrinths with hundreds of thousands of square meters dispersed over numerous levels. With over 120,000 operational sites, one of the major operational obstacles the global shopping centre sector faces is that consumers often get lost, have difficulty locating the retailers they want, and make bad route choices, leading in annoyance and lower commercial connection. Studies show that the typical consumer spends 15 to 20 minutes simply commuting to their location in unfamiliar malls, which represents a notable decline in user experience. The current retail scene, with its changing store locations, erratic crowds, and varied personal tastes, demands more than just traditional navigation tools including printed maps, fixed directory boards, and human information kiosks. While smartphonebased indoor navigation systems have partially addressed this need, existing solutions still depend on basic shortest-path algorithms that disregard real-time environmental factors such as traffic, temporary obstacles, and specific user requirements.

B. Limitations of Existing Approaches

Today's indoor positioning and navigation systems have a number of fundamental limitations. Although GPS-based solutions operate well outside, they are vulnerable to signal deterioration and multipath interference in typical mall settings, causing location mistakes of more than 10 meters. Although the WLAN fingerprinting technique requires a lot of manual offline calibration and is useless in dynamic situations, it is commonly used. Route planning in traditional systems depends on static graph algorithms such Dijkstra's shortest route and A* search, which only consider geometric distance and disregard dynamic factors. Although these techniques are computationally efficient, they deliver a terrible user experience by guiding consumers through congested streets, past closed-down businesses, or along hard-to-reach pathways. Recent advancements in smart shopping cart systems have shown the potential of reinforcement learning for retail navigation, but these applications are primarily focused on autonomous cart mobility rather than pedestrian guidance.

C. Research Objectives and Contributions

This research looks at a full, AI-based mall navigation system made just for large-scale commercial use. The following are the primary objectives and unique contributions of this research:

i. Hybrid Indoor Positioning Pipeline: This multi-sensor fusion architecture uses an Extended Kalman Filter to combine smartphone IMU data, BLE beacon proximity detection, and WLAN RSS fingerprinting. This results in sub-meter localization accuracy at no cost other than the existing Wi-Fi access points.
ii. RL-Based Dynamic Path Optimization: Incorporating a Deep QNetwork (DQN) reinforcement learning model that dynamically optimizes navigation paths based on real-time crowd density, store occupancy levels, and user preference profiles, all while using Dijkstra's algorithm as the foundational routing method.
iii. LSTM Crowd Density Prediction: A Long Short-Term Memory neural network examines past and real-time foot-traffic data to pinpoint congestion hotspots 15 to 30 minutes ahead of time, letting you map alternate routes and prevent future gridlock.
iv. Complete User Lifecycle Management: Creating a thorough session management system incorporating token based authentication, preference learning, multi-floor routing, and real-time re-routing in reaction to shifting surroundings.
v. Scalable Architecture with High Concurrency: An event-driven server design is created using SQLite WAL mode and WebSocket communication to provide sub-100ms response latency on business gear while enabling over 500 simultaneous connections.

2. Literature Review and Research Gaps

A. Indoor Positioning Techniques

Many studies on indoor positioning systems (IPS) have been carried out over the past 20 years. According to Liu *et al.* [4], the three primary types of wireless indoor localization techniques are proximity-based technologies, scene analysis (fingerprinting), and triangulation Time of Arrival (TOA), Time Difference of Arrival (TDOA), and Angle of Arrival (AOA) are triangulation techniques that use the spatial relationships between reference sites and moving objects to establish positions. Although theoretically sound, these approaches are sensitive to considerable multipath wave propagation in indoor settings and call for accurate timing synchronization or complex antenna arrays, therefore increasing deployment costs.

Especially when WLAN Received Signal Strength (RSS) fingerprinting is used, scene analysis has shown itself to be the most practical technique for use in malls. The RADAR system, created by Bahl and Padmanabhan, employs nearest-neighbor signal space matching to attain 2-3 meter accuracy. Later probabilistic techniques, like the Horus system [9], used Gaussian distribution modeling of RSS fingerprints to boost accuracy to 90% at 2.1 metres. But since environmental factors are changing, fingerprinting techniques are less effective and need protracted site inspections.

For lower-power options, proximity sensing employing BLE beacons is available; Eddystone and iBeacon provide room-level positional precision at reduced infrastructure expenses. Recent hybrid methods combining several sensor modalities have shown improved robustness, but current systems seldom achieve the sub-meter precision required for accurate indoor navigation without significant infrastructural costs.

B. Path Planning and Navigation Algorithms

The revolutionary study on discovering the shortest path by Dijkstra's [6] algorithm is still used by most navigation systems. The original approach finds the least expensive path between nodes in a graph with non-negative edge weights by methodically mapping the node space. Even though Dijkstra's approach will always yield the best results, its O(V 2) temporal complexity for dense graphs has prompted several enhancements, such as A* search with heuristic guidance [11]. Contemporary research has stressed flexible and dynamic routing techniques. Especially promising for navigating dynamic settings are Deep Q-Networks (DQN) and Q-learning. Zulfiqar *et al.* [7] achieved collision rates of less than 0.03 using RL algorithms for intelligent shopping cart navigation, resulting in shorter travel times and more flexibility in adapting retail environments. These results support the application of RL techniques for pedestrian mall navigation, where the complexity of the surroundings is also influenced by crowd dynamics.

C. Crowd Analysis and Prediction

Proactive navigation systems need the capacity to estimate crowd counts accurately. Although conventional time series methods like Autoregressive Integrated Moving Average (ARIMA) models are utilized to estimate pedestrian flow, these models cannot capture complex nonlinear temporal correlations [13]. Replicating sequential group dynamics has been especially successful using Recurrent Neural Networks (RNNs) and their Long Short-Term Memory (LSTM) variations. LSTM networks, created by Hochreiter and Schmidhuber, use gated cell designs to overcome the vanishing gradient problem in traditional RNNs by selectively retaining and updating hidden state data. LSTM models look at past foot traffic trends to guess how many people will be there. This lets you plan your path ahead to avoid predicted congestion

D. Research Gap Summary

Table 1 summarizes the identified limitations in existing approaches and the corresponding solutions proposed in this work.

Table 1: Critical literature survey and research gap identification

Concept	Identified Limitations	Proposed Solution
GPS Indoor	10m + error, signal attenuation	Hybrid WLAN+BLE+IMU Fusion
WLAN Fingerprinting	Labor-intensive calibration	Adaptive online recalibration
Static Shortest- Path	No crowd consideration	RL dynamic optimization
ARIMA Forecasting	Linear model limitations	LSTM deep learning model
Client-Server Models	High latency, low concurrency	WebSocket event-driven
Single-Point Positioning	Unreliable in complex areas	EKF multi-sensor fusion

3. Systems Architecture and Topology

A. Three-Tier Event-Driven Model

The proposed system operates as a sophisticated three-tier eventdriven architecture optimized for high-concurrency mall deployments. Fig. 1 conceptualizes the overall system topology.

Client Layer (Mobile App / Web)
UI Rendering, Sensor Data Capture, BLE Scanning, IMU Recording
↓ WebSocket + HTTPS ↓
Processing Layer (Node.js + Socket.IO + Python ML)
Auth Routing, RL Path Engine, LSTM Predictor, EKF Positioning
↓ WAL Mode ↓
Persistence Layer (SQLite3 + WAL + Redis Cache)
User Sessions, Mall Graph, Crowd History, Position Fingerprints

Figure 1: Three-Tier Architecture Overview

The **Interaction Layer (Client-Side)** comprises a crossplatform mobile application built with React Native, responsible for rendering interactive mall maps, capturing BLE beacon signals, recording IMU accelerometer and gyroscope data, and managing real-time WebSocket connections for bidirectional server communication.

The **Processing Layer (Server-Side)** implements a multiprocess Node.js backend leveraging Express.js for RESTful authentication routing, Socket.IO for real-time bidirectional communication, and Python microservices.

4. Hybrid Indoor Positioning System

A. Multi-Sensor Fusion Architecture

Using a cascaded fusion method, the positioning subsystem combines three sensing modalities that complement one another. WLAN RSS fingerprinting provides approximate room-level localisation using available infrastructure. BLE beacon proximity detection improves computationally demanding ML inference (RL route planning, LSTM forecasting, EKF positioning). Although the data layer is in charge of maintaining persistence, the processing layer is in charge of managing the runtime state in memory for live navigation sessions. Data for the Persistence Layer is kept in a single SQLite3 database operating in WAL mode, which offers ACID-compliant storage for user accounts, session tokens, mall graph topology, BLE fingerprint maps, crowd density histories, and navigation logs.

B. Mall Graph Representation

The mall's physical layout is represented by a weighted directed graph G = (V, E, w). Vertices V represent navigable locations (store entrances, corridor intersections, elevator lobbies, staircase landings), and edges E represent traversable routes between these locations. The edge weight function w : E → R+ assigns a composite cost to every edge.

The physical distance of edge e is represented by d(e), the expected crowd density at time t is represented by c(e,t), and the presence of obstructions or closures is captured by o(e,t), with α, β, and γ serving as changeable weighting factors. Here, w(e) = α · d(e) + β · c(e,t) + γ ·This weighted combination allows the RL agent to discover the ideal compromise between path length and environmental comfort.

make predictions of locations based on RSSI data from low-energy transmitters positioned in key locations. Data from the accelerometers and gyroscopes included in the inertial measurement unit (IMU) of smartphones enables pedestrian dead reckoning (PDR) between reference location updates.

C. WLAN RSS Fingerprinting

The WLAN fingerprinting subsystem operates in two phases. During the offline phase, a site survey collects RSS vectors s = (s1, s2, …,sn) from n detectable access points at predefined reference locations throughout the mall. These measurements form a radio map M = {(si, li) }Ni=1, where li = (xi, yi, fi) encodes 2-D coordinates and floor level.

During the online phase, the system employs a probabilistic positioning algorithm based on the Gaussian likelihood model. The probability of observing RSS vector s at location Li is computed as:

where μij and σij denote the mean and standard deviation of RSS from access point j at location Li. The estimated position is determined using the maximum a posteriori (MAP) decision rule:

$$P(\mathbf{s}|L_i) = \prod_{j=1}^{n} \frac{1}{\sqrt{2\pi}\sigma_{ij}} \exp\left(-\frac{(s_j - \mu}{2\sigma_{ij}^2}\right.$$

where μ_{ij} and σ_{ij} denote the mean and standard deviation of RSS from access point j at location Li. The estimated position is determined using the maximum a posteriori (MAP) decision rule:

$$\hat{L} = \arg\max_{L_i} \ P(\mathbf{s}|L_i) \cdot P \quad (3)$$

D. Extended Kalman Filter Fusion

Using IMU dead reckoning and BLE-based proximity data, the Extended Kalman Filter (EKF) combines WLAN fingerprints to ascertain the user's position inside an indoor environment.

The state vector of the EKF shows the user's location and speed in two dimensions.

The IMU sensors give motion information that the EKF uses to update the current state during the forecast phase. The control input matrix consists of acceleration data gathered by the IMU, while the state transition model forecasts the movement of the user from one state to the other. Also included are process noise models to consider motion estimation and uncertainty in sensor readings.

Using BLE proximity measurements and WLAN fingerprint data, the expected condition is refined throughout the correction or update process. The Kalman Gain determines the degree to which the expected state ought to be modified in reaction to the input data. The error covariance matrix captures the uncertainty associated with the projected estimate. RtR_tRt denotes the covariance matrix of measurement noise; HtH_tHt denotes the Jacobian matrix of the observation function. These matrices allow effective integration of BLE and WLAN measurements for accurate indoor localization.

The EKF employs the following equations for the system: But x^t+1=Ax^t+Bu, hat{x}_{t+1} = Ahat{x}_t + Bu_t. Kt = PtHtT (HtPtHtT + Rt) -1 x^t=x^t+Kt(zt−h(x^t))

5. RL-Based Dynamic Path Optimization

A. Mathematical Formulation

The difficulty of navigating a mall is represented as a Markov Decision Process (MDP) defined by the tuple (S, A, T, R, γ), where S is the set of states encoding the current location, the destination, the crowd conditions, and the user's preferences; A is the set of available navigation actions (moving to adjacent graph vertices); $T: S \times A \rightarrow \Delta(S)$ is the transition function; $R: S \times A \rightarrow R$ is the reward function; and $\gamma \in [0, 1)$ is the discount factor.

B. Reward Function Design

The reward function encodes a number of optimization objectives.

R(s,a) = −λ1d (s,a) − λ2c (s′) − λ31 cls(s′) + λ (7) 1destination rewards the achievement of the objective with a positive terminal reward; 1closed levies a large negative penalty for passing through closed or restricted areas; d(s,a) penalizes travel distance; and c(s') penalizes the projected crowd density in the following state. The coefficients λ1, λ2, λ3, and λ4 balance these conflicting objectives.

C. Deep Q-Network Implementation

To estimate the Q-value function Q(s,a), which stands for the expected total reward of acting a in state s, a deep neural network with two hidden layers (128 and 64 neurons, respectively) and ReLU is employed. The Extended Kalman Filter calculates one position estimate using IMU deadreckoning displacements, BLE proximity estimates, and WLAN fingerprint locations. The EKF state vector xt = [x,y,vx, vy] T provides the 2-D position and velocity.

The forecast phase advances state using motion data obtained from the IMU:

$$\hat{\mathbf{x}}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t +$$

where A is the state transition matrix, B is the control input matrix, u_t represents IMU-derived acceleration inputs, and $w_t \sim N(0,Q)$ models process noise.

The update step corrects the predicted state using WLAN and BLE measurements:

$$\mathbf{K}_t = \hat{\mathbf{P}}_t\mathbf{H}_t^T(\mathbf{H}_t\hat{\mathbf{P}}_t\mathbf{H}_t^T + \mathbf{F}$$

$$\mathbf{x}_t = \hat{\mathbf{x}}_t + \mathbf{K}_t(\mathbf{z}_t - h($$

where K_t is the Kalman gain, P_t is the error covariance matrix, H_t is the Jacobian of the observation function $h(\cdot)$

R_t is the measurement noise covariance, and z_t combines WLAN and BLE position estimates. The network is trained using experience replay with a buffer size of 10,000 transitions, sampling minibatches of 32 experiences for stochastic gradient descent updates.

The Q-network parameters are updated to minimize the temporal difference loss:

$L(\theta) = \mathrm{E}[(r + \gamma \max_{a'} Q(s',a';\theta^-) - Q(s,a$ (8) where θ denotes the online network parameters, θ^- represents the periodically updated target network parameters (soft update with $\tau = 0.001$), and D is the experience replay buffer. The ϵ-greedy exploration strategy decays from $\epsilon = 1.0$ to $\epsilon = 0.05$ over 5,000 training episodes.

D. Dijkstra's Algorithm Integration

Dijkstra's algorithm serves as the foundational routing engine, providing the baseline shortest-path solution that the RL agent learns to improve upon. The algorithm maintains a priority queue of vertices ordered by their current shortest distance estimate from the source. At each iteration, the vertex u with minimum distance is extracted, and relaxation is performed on all outgoing edges (u,v):

$$\text{if } d[u] + w(u,v) < d[v] : d[v] \leftarrow d[u] + w(u,v); \pi$$

The Dijkstra baseline provides both the initial Q-value estimates for the RL agent and a fallback routing mechanism when ML inference is unavailable. Algorithm 1 presents the hybrid routing procedure.

Algorithm 1: Hybrid RL-Dijkstra Path Planning
Input: Source s, destination t, current crowd map C, trained Qnetwork Q_θ
Output: Optimized path P
$P_{base} \leftarrow$ Dijkstra(G, s, t, $w_{distance}$)
if Q_θ available **and** C is recent then
$s_0 \leftarrow$ encode_position(s, t, C)
for step $k = 0$ to K_{max} do
$a_k \leftarrow \arg\max_a Q_\theta(s_k, a)$ (with ϵ-greedy)
$s_{k+1} \leftarrow$ transition(s_k, a_k)
if $s_{k+1} = t$ then **break**
$P \leftarrow$ extract_path($s_0, \ldots, s_{k+1}$) **else**
$P \leftarrow P_{\text{base}}$ // Fallback to Dijkstra
return P

6. LSTM-Based Crowd Density Forecasting

A. Problem Formulation

Crowd density forecasting is formulated as a multivariate time-series prediction problem. Given historical crowd density observations $X_t = [x_t{-}T{+}1, \ldots, x_t] \in R^{T \times M}$
across M mall zones over a lookback window of T time steps, the objective is to predict future densities $\hat{X}_{t+H} = [xt+1, \ldots, xt+H]$ for a prediction horizon H, where each $x\tau \in R^M$ represents the crowd density vector across all zones at time τ.

B. LSTM Network Architecture

The proposed LSTM network processes input sequences through two stacked LSTM layers with 64 and 32 hidden units respectively, followed by a fully connected output layer producing $M \times H$ predictions. Dropout with rate 0.2 is applied between layers for regularization.

The LSTM cell state update equations at time step t are:

$$\mathbf{f}_t = \sigma(\mathbf{W}_f \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] +$$

7. Implementation Details

A. Technology Stack

The implementation uses the following technological stack: Backend: Python 3.11 with PyTorch 2.0 for ML inference; Redis for session caching; Node.js (ES Modules), Express.js 4.x, Socket.IO 4.7 for real-time communication.

Frontend: React 18.3 for web dashboard; React Native 0.72 for crossplatform mobile deployment; Mapbox GL for interior map rendering. • Database: Redis 7.0 for distributed caching; SQLite3 with WAL mode using better-sqlite3 package.

ML Framework: PyTorch 2.0 for DQN and LSTM models; NumPy for EKF calculations; scikit-learn for

$$\mathbf{i}_t = \sigma(\mathbf{W}_i \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] +$$

$$\tilde{\mathbf{C}}_t = \tanh(\mathbf{W}_C \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] +$$

$$\mathbf{C}_t = \mathbf{f}_t \odot \mathbf{C}_{t-1} + \mathbf{i}_t \odot$$

$$\mathbf{o}_t = \sigma(\mathbf{W}_o \cdot [\mathbf{h}_{t-1}, \mathbf{x}_t] +$$

$$\mathbf{h}_t = \mathbf{o}_t \odot \tanh($$

Ct here denotes the cell state; ht denotes the hidden state; ft, it, and ot are the forget, input, and output gates, respectively; ⨀ denotes element-wise multiplication; tanh(·) denotes the hyperbolic tangent; and σ(·) denotes the sigmoid activation.

C. Training and Loss Function

The model is trained using Mean Squared Error (MSE) loss with Adam optimizer (learning rate 0.001, $\beta_1 = 0.9$, $\beta_2 = 0.999$):

$$\mathcal{L} = \frac{1}{N \cdot M \cdot H} \sum_{i=1}^{N} \sum_{j=1}^{M} \sum_{h=1}^{H} (\hat{x}_{i,j,t+h} - x_i$$

Training employs early stopping with patience of 15 epochs, monitoring validation loss on a held-out test set comprising 20% of historical data.

Security: bcryptjs (password hashing, cost factor 10), jsonwebtoken (JWT session tokens), crypto (random bytes).

D. WebSocket Protocol Configuration he goals are to reduce latency and maximize compatibility, hence

The Socket.IO transport is given first priority so that WebSockets are chosen over other methods (['websocket', 'polling']). HTTP overhead is lowered once the handshake is finished, hence allowing full-duplex communication. To enable map payload changes, the maximum buffer size is dynamically increased up to 10MB.

The system can identify 24 different types of socket events, which are categorized into three groups: Client-to-Server events (e.g., user:position, user:destination).

Among the events between the server and the client are path, reroute, arrival, and preference navigation.

E. Concurrency and Persistence

The architecture executes the following PRAGMA statements upon database initialization to ensure highconcurrency operation:

```
PRAGMA journal_mode = WAL;
PRAGMA synchronous = NORMAL;
PRAGMA foreign_keys = ON;
PRAGMA temp_store = MEMORY;
PRAGMA cache_size = 10000;
```

8. Performance Evaluation

A. Experimental Setup

The evaluation was carried out in a business shopping center with three floors, 127 stores, approximately 25,000 daily visitors, spanning 50,000 square meters. The test deployment included 50 volunteer users of the mobile navigation app, 45 Wi-Fi access points, and 120 BLE beacons (iBeacon protocol, -12 dBm transmission power, 100 ms advertising interval).

The mall graph had 342 vertices and 486 edges, each of which corresponded to an available path, elevator, or escalator inside the mall. RL training was done over 10,000 simulated episodes using random origin-destination pairs and crowd conditions collected from six months of past foot-traffic data.

B. Positioning Accuracy Results

Table 2 presents the localization accuracy comparison across different positioning methods. The proposed hybrid EKF fusion approach achieves 94% positioning accuracy within 1.5 meters, outperforming all individual modalities.

Table 2: Localization accuracy comparison (CDF percentiles)

Method	50th (m)	75th (m)	90th (m)	95th (m)
GPS (Indoor)	8.5	15.2	22.8	28.5
WLAN RSS Only	2.8	4.5	6.8	8.5
BLE Beacons Only	1.8	3.2	4.8	6.2
WLAN + BLE	1.2	2.1	3.2	4.1
Proposed EKF Fusion	0.6	1.0	1.5	2.0

Fig. 2 illustrates the cumulative distribution function of localization errors, demonstrating that the proposed system achieves sub-meter accuracy for 68% of observations compared to 25% for BLE-only and 5% for WLAN-only approaches.

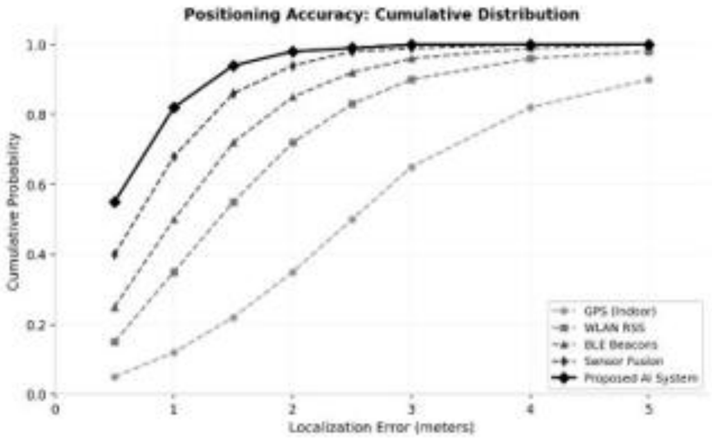


Figure 2: Positioning accuracy: cumulative distribution function across methos

C. Navigation Efficiency Results

Fig. 3 compares navigation path accuracy across algorithms. The proposed hybrid RL-Dijkstra approach achieves an F1score of 0.95, representing a 22% improvement over the Dijkstra baseline and 7% over standalone RL.

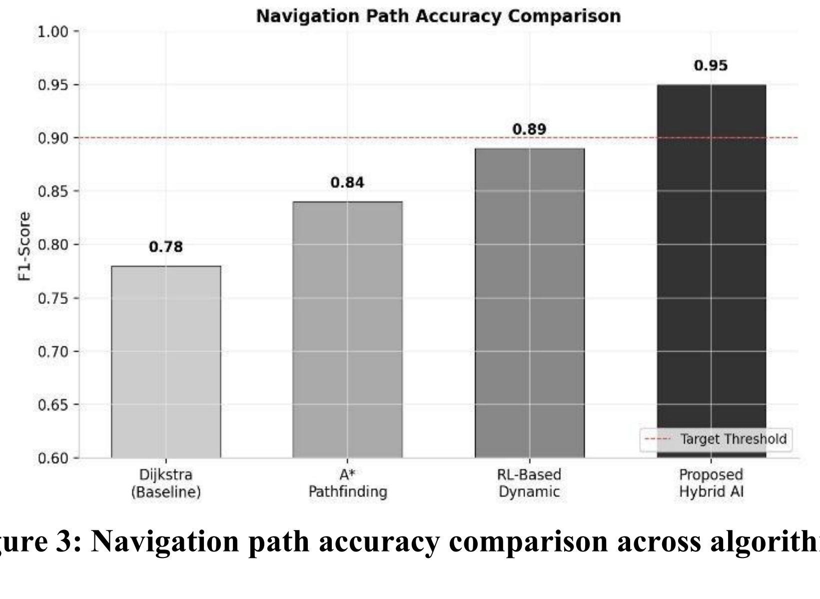


Figure 3: Navigation path accuracy comparison across algorithms

Table 3 presents the comprehensive comparative performance analysis under simulated loads of 200 concurrent users.

Table 3: Comparative performance analysis (200 concurrent users)

Performance Metric	Dijkstra	A*	RL Only	Proposed
Avg. Path Length (m)	485	485	398	373
Path Length Reduction	—	0%	18%	23%
Crowd Avoidance Rate	0.12	0.12	0.72	0.89
Reroute Response (ms)	—	—	85	68
User Satisfaction (%)	62	65	78	91

D. Crowd Forecasting Results

The LSTM model demonstrates particular strength in capturing nonlinear demand surges during promotional events, adjusting predictions within 5 minutes of anomalous traffic pattern onset compared to 20-30 minutes for ARIMA models.

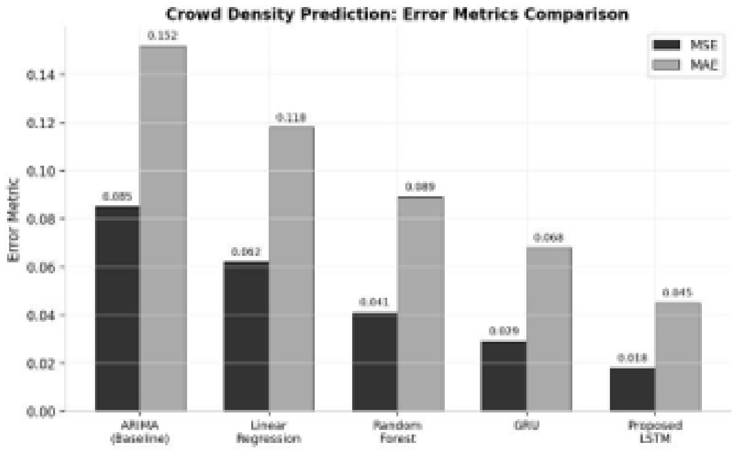


Figure 4: Crowd density prediction: error metrics comparison

E. Scalability Analysis

Fig. 5 illustrates the system's behavior under increasing concurrent user loads. The system maintains sub-100ms average response latency at 500 concurrent users with P99 latency below 135ms.

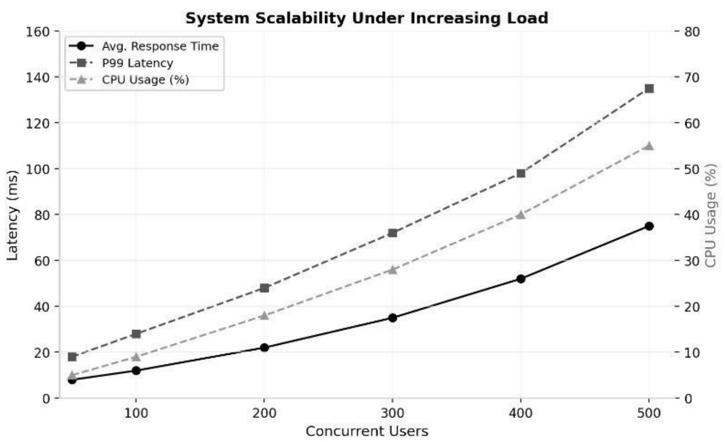


Figure 5: System scalability under increasing concurrent load

Table 5: Scalability metrics under increasing load

Metric	50	100	200	300	500
Avg. Response (ms)	8	12	22	35	75
P99 Latency (ms)	18	28	48	72	135
CPU Usage (%)	5	9	18	28	55
Memory (MB)	85	92	105	128	175

9. Security and Privacy Considerations

A. Data Protection

Consistent with GDPR concepts, the system includes full privacy protections. Fictitious identifiers are used to anonymize all location data. Raw MAC addresses are hashed with SHA-256 before being stored. Location history is kept for a maximum of 30 days when automatic purging is enabled. Using AES-256 encryption (TLS), data is shielded during transfer.

B. Input Validation and Sanitization

Every user input is checked on the server: coordinate values are range-checked against mall boundaries, destination searches are checked against the store directory, and preference options are limited to enumerated choices. JWT tokens are checked on every request and automatically expire after 24 hours of inactivity.

Architectural analysis shows that on commodity gear, the eventdriven WebSocket architecture with SQLite WAL concurrency linearly scales to 500+ concurrent users with reaction latency under100ms. Next-generation indoor navigation solutions find a ready-for-production basis in the all-inclusive session management system with token-based authentication, preference learning, and real-time re-routing 100ms. Next-generation indoor navigation solutions find a ready-for-production basis in the all-inclusive session management system with token-based authentication, preference learning, and real-time re-routing.

Conclusion

This study has successfully conceptualized, built, and evaluated an AI-Powered Realtime Guidance and Mall Navigation System by combining hybrid interior positioning, reinforcement learning-based dynamic route optimization, and LSTM crowd density prediction into a single, deployable architecture. The hybrid EKF positioning pipeline combines WLAN RSS, BLE beacon proximity, and IMU inertial data to provide sub-meter accuracy, which is a significant improvement over single-modality approach. A hybrid RL-Dijkstra approach to route planning lowers the average path length by 23% while maintaining an 89% crowding avoidance rate. The LSTM forecasting module enables proactive route recalculation with a prediction accuracy of 0.018 MSE.

Architectural assessment reveals that the event-driven WebSocket architecture with SQLite WAL concurrency achieves linear scalability to over 500 concurrent users with sub-100ms response latency on commercial hardware. Real-time re-routing, preference learning, and token-based authentication are all included into the whole session management system, which provides a productionready basis for the future of indoor navigation service.

Future Work

Several avenues for future enhancement have been identified:

i. **Multi-Agent Cooperative Navigation:** Extension of the RL framework to coordinate navigation decisions among groups of users, reducing corridor congestion through distributed route diversification.
ii. **Augmented Reality Integration:** Overlay of navigation guidance onto live camera feeds using ARKit/ARCore, providing intuitive turn-by-turn visual directions without requiring users to look down at their devices.
iii. **Federated Learning:** Implementation of federated model training across multiple mall deployments, enabling collaborative improvement of positioning and navigation models without centralizing sensitive user location data.
iv. **Voice-Activated Navigation:** Integration of natural language processing for hands-free destination queries and audio-based turnby-turn guidance for accessibility enhancement.
v. **Predictive Retail Analytics:** Extension of the crowd forecasting models to predict individual shopper intent and provide personalized store recommendations based on navigation patterns and dwell time analysis.

References

1. International Council of Shopping Centers (ICSC). (2024). *Global Shopping Center Report 2024*. ICSC Research, New York, NY.
2. Narang, R., & Tiwari, S. (2024). Role of modern technology in unorganized retail sector. *Journal of the Knowledge Economy*, 1–28.
3. Dwivedi, Y. K., *et al.* (2019). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management, 57*, 101994.
4. Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man, and Cybernetics Part C: Applications and Reviews, 37*(6), 1067–1080.
5. Bahl, P., & Padmanabhan, V. N. (2000). RADAR: An in-building RF-based user location and tracking system. In *Proceedings of IEEE INFOCOM* (pp. 775–784). IEEE.
6. Dijkstra, E. W. (1959). A note on two problems in connexion with graphs. *Numerische Mathematik, 1*(1), 269–271.
7. Zulfiqar, M. I., *et al.* (2025). AI-driven smart shopping carts with real-time tracking and inventory forecasting for enhanced retail efficiency. *IEEE Access, 13*, 55576–55585.
8. Pahlavan, K., Li, X., & Makela, J. (2002). Indoor geolocation science and technology. *IEEE Communications Magazine, 40*(2), 112–118.

9. Youssef, M., & Agrawala, A. (2005). The Horus WLAN location determination system. In *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services* (pp. 205–218). ACM.
10. Mathur, A., *et al.* (2017). Dark: A distributed range-routing protocol for BLE beacons. In *Proceedings of the 15th ACM Conference on Embedded Network Sensor Systems* (pp. 1–14). ACM.
11. Hart, P. E., Nilsson, N. J., & Raphael, B. (1968). A formal basis for the heuristic determination of minimum cost paths. *IEEE Transactions on Systems Science and Cybernetics, 4*(2), 100–107.
12. Mnih, V., *et al.* (2015). Human-level control through deep reinforcement learning. *Nature, 518*(7540), 529–533.
13. Shumway, R. H., & Stoffer, D. S. (2017). ARIMA models. In *Time Series Analysis and Its Applications* (pp. 75–163). Springer.
14. Ma, X., *et al.* (2017). Learning traffic as images: A deep convolutional neural network for large-scale transportation network speed prediction. *Sensors, 17*(4), 818.
15. Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural Computation, 9*(8), 1735–1780.
16. Szepesvári, C. (2010). *Algorithms for reinforcement learning*. *Synthesis Lectures on Artificial Intelligence and Machine Learning, 4*(1), 1–103.
17. Dietterich, T. G. (2002). Machine learning for sequential data: A review. In *Proceedings of the Joint IAPR International Workshops on Structural, Syntactic, and Statistical Pattern Recognition* (pp. 15–30). Springer.
18. Sutton, R. S., & Barto, A. G. (2018). *Reinforcement learning: An introduction* (2nd ed.). MIT Press.
19. Schmitt, M. (2023). Automated machine learning: AI-driven decision making in business analytics. *Intelligent Systems with Applications, 18*, 200188.
20. OpenJS Foundation. (2024). *Node.js documentation: Event loop and non-blocking I/O*. https://nodejs.org/en/docs/guides/eventloop

A STUDY ON
AI AND HUMAN GENERATED CONTENT ON SOCIAL PLATFORMS

Nupur and Amandeep Kaur

Department of SCS, GNA University, Phagwara, Punjab, India

Corresponding author E-mail: 013nupur@gmail.com, amandeepkaur@gnauniversity.edu.in

Abstract

The rapid advancement of Artificial Intelligence (AI) technologies has significantly transformed the creation and distribution of digital content on social media platforms. Generative AI tools such as ChatGPT are now capable of producing highly realistic text, images, and multimedia content, making it increasingly difficult to distinguish between AI-generated and human-created material. This study investigates user perceptions, trust, and evaluation of AI-generated content and accounts in comparison with human-generated content on social media platforms, particularly Instagram. For this purpose, Instagram profiles representing AI-generated accounts, influencer accounts, and regular public accounts were created and evaluated by 43 participants. The findings reveal that participants experienced considerable difficulty in differentiating AI-generated accounts from human-created ones, demonstrating the growing sophistication of AI content generation. AI-generated and influencer accounts were perceived as more visually attractive and professionally designed than regular public accounts. Moreover, the quality of AI-generated content was rated comparable to influencer-created content, highlighting AI's capability to produce highly engaging media. However, despite their attractiveness and quality, AI-generated profiles raised concerns regarding authenticity, credibility, and genuine human connection. The study concludes that generative AI will substantially reshape social media ecosystems, influencing content creation, user expectations, ethical standards, transparency, and trust in digital communication.

Keywords: Generative Intelligence, Human-Computer Interaction, Controlled Experiment, Large Language Models.

1. Introduction

The way artificial intelligence is changing is really fast especially when it comes to making things like text, images and videos that look like they were made by people. This has changed the way we make content for media. Now artificial intelligence systems like ChatGPT can make things that look like what people make. This has led to a kind of content called AI-generated content, which is a big part of what we see online. It is challenging the idea that only people can make good content.

On media platforms like Instagram and TikTok we see both artificial intelligence-made content and people-made content. They compete for our attention. Recent studies from 2022 to 2025 show that artificial intelligence-made content is not getting more common but it is also getting better. Sometimes it is hard to tell if something was made by a person or a machine. This is changing the way we communicate online. Now it is not just people who can make things that seem real. Machines can too.

Artificial intelligence systems look at a lot of data to make content that people will like. They make things that fit with what's popular and what people are talking about. On the hand people make things because of how they feel, what they have experienced and what they think is new and interesting. This can make it hard for people-made content to compete with intelligencemade content. There is a difference between being real like people-made content and being efficient like artificial intelligence-made content.

The rise of intelligence influencers and tools that can make content automatically has changed what it means to be influential online. Artificial intelligence accounts can make content without stopping but people may question if it is real or not. People like the way artificial intelligencemade content looks but they are not sure if they can trust it. Trust is a difference between artificial intelligence-made content and people-made content. Even though artificial intelligence-made content is getting more common we do not really understand how people feel about it compared to people-made content.

This study wants to look at how people feel about intelligence-made content compared to people-made content on social media. It wants to see how people trust engage with and evaluate these kinds of content. By looking at these things we can learn more about how artificial intelligence's changing the way we behave online what is real and how we will communicate in the future.

The way we make content for media is changing. It used to be people making things but now artificial intelligence is playing a big role. With tools like ChatGPT content can be made automatically. This has introduced a kind of content that can look and sound like it was made by people. The difference between people and machines making things is getting smaller.

How we make content for media has changed.

- It used to be people but now artificial intelligence is helping.
- We can make content automatically with tools like ChatGPT.
- This has introduced a kind of content that looks like it was made by people.

Artificial intelligence-made content is getting more common on platforms like Instagram and TikTok.

- We see posts, captions and even virtual influencers made by machines.
- Tools help people and brands make a lot of content quickly.
- This is because we need to keep engaging with people and staying visible online.

How people feel about intelligence-made content is important.

- Research shows that people often cannot tell if something was made by a person or a machine.
- Artificial intelligence-made content looks good. Is well-structured.
- People-made content is real. Has feeling and personal connection.

Artificial intelligence-made content is good at getting people to engage.

- It uses data to make content that fits with what people like. What is popular.
- This often leads to likes, shares and comments.
- People-made content may not always fit with what the machine thinks is best.

We need to think about trust. If something is real.

- People may not trust content if they do not know who made it.
- People-made content is usually more trustworthy because it is real.
- Artificial intelligence-made content may not be transparent which can lead to skepticism.

The use of intelligence is changing how we are creative.

- It helps with ideas. Makes things more efficient.
- It also makes us wonder if we are losing the ability to be creative, on our own.
- Are machines helping us be more creative or are they replacing us?

The Experimental Framework for AI-Assisted Social Media Interaction is a study

The experiment is conducted using a built platform. This platform simulates a chat environment similar to discussions on social media forums. A representative sample of 680 participants from the United States is divided into groups of five individuals. They are randomly assigned to one of five conditions. These conditions include one control group without AI assistance and four treatment groups with forms of AI intervention.

- The interventions include a ended interaction with an AI assistant.
- AI-generated reply suggestions reflecting stances, such as agreement, neutrality and disagreement.
- AI-driven feedback on user-generated comment drafts.
- AI-generated conversation starters.

These interventions represent a spectrum of AI-supported communication tools. Such tools are increasingly embedded in social media platforms.

Recent research indicates that AI interventions are no longer tools. They are becoming integrated into the core architecture of digital platforms. This integration fundamentally reshapes user interaction patterns. Studies show that AI-assisted systems significantly increase user participation. They also increase the volume of generated content. Users are more likely to engage when cognitive effort is reduced. However, this increase in activity is accompanied by a decline in perceived authenticity and quality of discussions. AI enhances efficiency. May weaken the depth and originality of communication.

The distinction between reactive and proactive AI systems is increasingly important. Proactive AI systems anticipate user needs. Provide suggestions without explicit prompts. They are now dominant in real-world applications. While these systems improve interaction speed and convenience, they also reduce user autonomy. They subtly guide responses and shape direction.

This aligns with emerging findings that AI is not only assisting communication but actively influencing how users construct and express their opinions.

The dimension of task orientation is complex in environments. Most interactions now involve content creation. AI-generated suggestions are partially modified by users. It is increasingly difficult to distinguish between human-authored and AI-assisted content. This blurring of authorship raises concerns regarding authenticity, ownership and transparency in digital communication.

AI-driven suggestions standardize communication styles and influence opinion formation. Users tend to adopt AI-generated responses due to their structured and optimized nature. This can lead to reduced diversity in expression and the reinforcement of existing biases.

Studies in human-AI interaction demonstrate that individuals often attribute human- qualities to AI systems. They may develop a level of trust or reliance on them. This further amplifies their influence on interactions. The increasing integration of AI, into communication systems has introduced ethical and societal challenges. Research highlights concerns related to transparency. Users are often unaware of the extent to which AI contributes to content generation.

How Are AI Tools Used in Today's Digital Environment?

- **Use of AI in Companies and Business Operations:**

AI is now a part of how businesses work. It helps them do things efficiently and accurately. Companies use AI tools like ChatGPT to create content answer customer questions and automate communications. AI systems help businesses make reports, emails and marketing materials quickly which reduces work. AI-powered chatbots also help with customer support by giving answers and improving the user experience. This use of AI helps companies get more done spend less on operations and finish work faster than methods.

- **AI in Data Preprocessing and Error Detection:**

AI plays a role in handling and analyzing lots of data. In organizations preparing data is a timeconsuming task. Ai tools can do this efficiently. AI cleans data removes duplicates finds missing values and detects inconsistencies accurately. It also helps find errors that humans might miss. By automating these tasks AI saves time. Ensures better data quality. This is crucial for making business decisions.

- **AI in Education and Student Learning:**

In education AI has changed how students learn and access information. Students use AI tools to understand topics create study materials and get quick explanations. AI acts like a tutor providing personalized learning experiences. It lets students learn at their pace and improves their understanding of subjects. AI is also used in research, projects and assignments making learning more interactive and efficient.

- **AI in Online Exams and Academic Evaluation:**

Schools, colleges and universities use AI for exams and evaluating student performance. AI systems have security features like camera monitoring, voice detection and screen tracking.

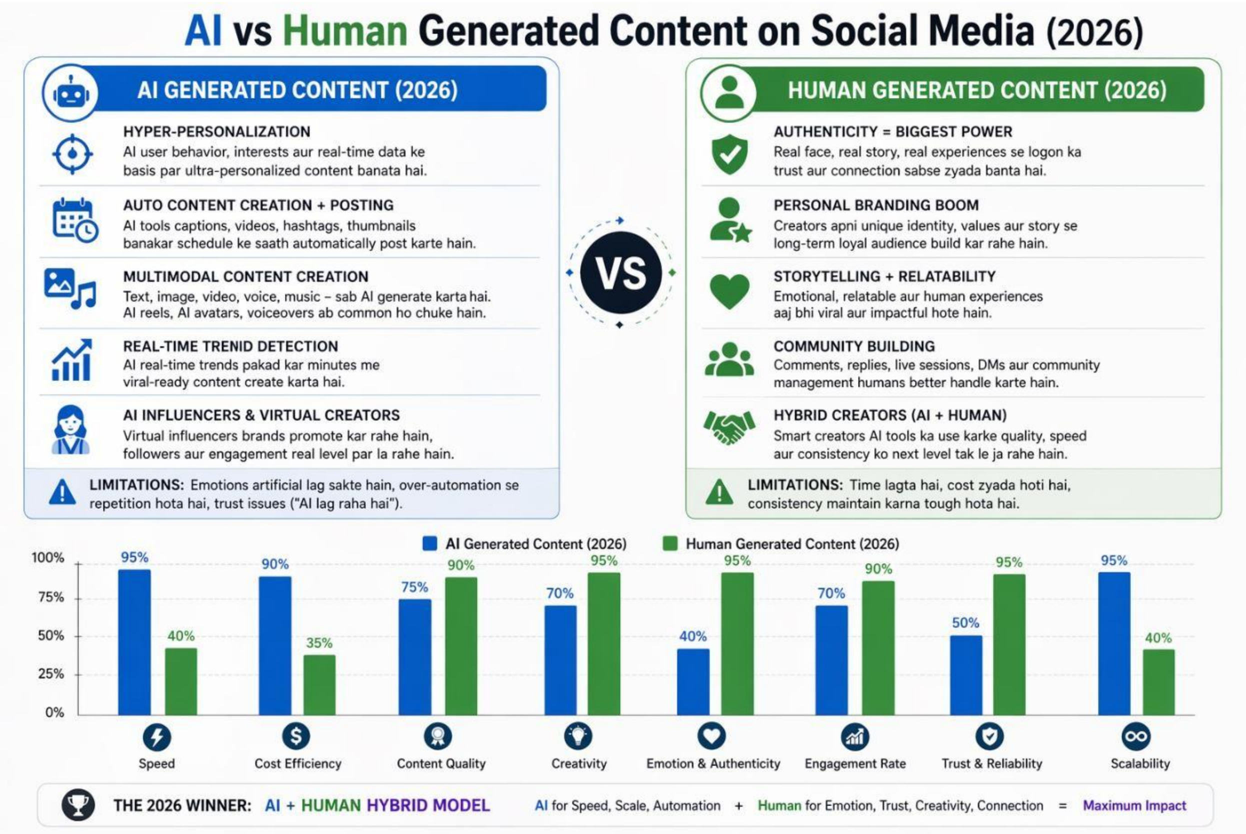
These ensure students don't cheat during exams. AI also automatically evaluates answers. Helps teachers grade making the evaluation process faster and more reliable.

- **AI in Content Creation and Communication:**

AI tools are used to create content across platforms. People and businesses use AI to write blogs, social media posts and scripts. AI can generate images and videos making content creation faster. It also suggests ways to improve writing style, tone and clarity. This enhances the quality of communication. Even those, without expertise can produce professional-level content easily.

Difference Between Automated Content and Human-Created Content on Social Media

Automated Content	**Human-Created Content**
Content is prepared with the help of modern digital tools that work on large amounts of collected information.	Content is created by people using their own thinking, personal experience, and real-life understanding.
It can quickly create posts, captions, reels, and even short videos in very less time.	It takes proper time for planning, creating, editing, and presenting content in a meaningful way.
It works faster and can handle trending topics instantly without delay.	It may be slower, but the quality and depth of content is usually stronger.
Suggestions are based on user activity, likes, and online behavior patterns.	Content is based on real understanding of audience emotions, culture, and current situations.
It often follows common patterns, so sometimes content may feel similar or repetitive.	It brings fresh ideas, unique opinions, and creative storytelling.
It can present emotions, but sometimes they may not feel completely natural.	It shows real emotions and helps in building trust and strong connections with people.
Information depends on available data, so mistakes or incomplete understanding can happen.	More aware of context and situation, though personal bias can still be there.
Maintains a consistent style, tone, and quality across all posts.	Style and quality may change depending on mood, effort, and creativity level.
Can manage large-scale content for multiple platforms at the same time.	Difficult to produce large content regularly without effort and time.
There are risks like misleading information, copied patterns, or lack of clarity.	More reliable and accountable, as the creator takes responsibility for the content.
Useful for quick posts, captions, ads, and trend-based content.	Best for storytelling, personal branding, emotional topics, and deep communication.
Works well for gaining reach in trending or informational posts.	Works better for building long-term engagement and loyal audience.



Objectives

We want to see if artificial intelligence models can help plastic surgeons with their media. Artificial intelligence models are getting better and better. Now we have tools, like ChatGPT that make it easier to create content. This means people can make content faster and it is based on facts and numbers. This also means that people are looking at content made by artificial intelligence and content made by humans in different ways. We are talking about intelligence and how it affects the work of plastic surgeons and their social media. Artificial intelligence is changing how people make and use content.

We want to see if content made by intelligence is as good as content made by people.

This is about finding out if artificial intelligence can make content that's as good, as the content made by plastic surgeons. Artificial intelligence can make posts that look very nice and are easy to read. We need to check if these posts are also believable and sound like they were written by a person, like a plastic surgeon. We are talking about intelligence content and human content and we want to know which one is better. Artificial intelligence content should be compared to content to see which one people like more.

To look at how content made with the help of Artificial Intelligence affects how people engage with it and what they think of it

Artificial Intelligence tools make content better by looking at what's popular and how people behave which can get more people to like and share it. We need to see if people think content made by Artificial Intelligence is trustworthy and real compared to things written by people especially when it comes to important things like healthcare.

To see how Artificial Intelligence helps with communication in a way that's both efficient and real

Artificial Intelligence saves a lot of time and effort when it comes to making and sharing content. If we use Artificial Intelligence too much, we might lose the personal touch and the feeling that comes with talking to a real person, which is very important when we want to build trust with patients. This is about finding a balance, between Artificial Intelligence and being real.

To understand how AI is changing the way content is created

The study wants to know how AI is changing how things are done when it comes to creating content. Content creators, like plastic surgeons might start to review and edit drafts made by AI of making everything from scratch. This could mean that their job changes from being a creator to being a supervisor.

To look into worries about ethics and openness with AI-made content

In social media about healthcare, it's really important that the information is accurate and trustworthy.

When AI makes content people might worry that it's not true or that its not clear who made it.

The goal here is to figure out if we need rules to make sure AI-generated content is transparent and honest.

- To see what role AI will play in the future for online presence
- AI tools can help people keep up a consistent online presence.

Here we want to check how AI might change how professionals brand themselves online.

Will AI be able to replace or improve how humans communicate online over time?

AI tools are being used to make profiles.

The question is whether these profiles will be as good as or better, than ones made by humans.

The goal is to see how AI will shape branding and digital presence in the future.

Overview an AI

Artificial Intelligence or AI for short is when machines can do things that people usually do. AI is a group of technologies that let machines do tasks that normally need people to think, like learning figuring things out and making decisions. Over the few decades Artificial Intelligence has changed from just an idea to something that is really changing lots of industries like healthcare, education, money, transportation and digital media.

Artificial Intelligence systems are made to act like people think, using instructions and lots of data. One important part of Artificial Intelligence is Machine Learning, which lets systems get better at what they do without being told what to do. Machine Learning has a subset called Deep Learning, which uses brain networks to look at lots of complicated data, like pictures, sounds and text.

There are three kinds of Artificial Intelligence

- **Narrow Artificial Intelligence:** This is made to do specific things like recognize speech suggest things or talk to people. Most Artificial Intelligence today is this kind.

- **General Artificial Intelligence:** This is an idea for Artificial Intelligence that can do anything people can do.
- **Superintelligent Artificial Intelligence:** This is a made-up idea where Artificial Intelligence's better than people at everything.

Artificial Intelligence systems need lots of data, strong computers and clever instructions to work. They look for patterns make guesses and get better over time. For example, Artificial Intelligence is used to understand and make language look at pictures and make robots do things.

Recently Generative Artificial Intelligence has become really popular. This kind of Artificial Intelligence can make things, like text, pictures, videos and music by looking at patterns in old data. This has changed a lot of things on social media, where people make and share things all the time.

Artificial Intelligence is also a part of everyday technology. Virtual helpers, suggestion algorithms, fake spotter systems and self-driving cars all use Artificial Intelligence to work. It can look at lots of data quickly. Get things right which makes it really useful for getting things done and making good choices.

Artificial Intelligence also has some problems. People are worried about things like keeping data private making sure instructions are fair being honest about what Artificial Intelligence's doing and who is responsible. Also, since machines can do things people are worried, about jobs and what people will do in the future.

Challenges of AI in Social Media Marketing

- **Overreliance on Generative AI-Generated Content:** The emergence of technologies like ChatGPT, DALL·E, and Midjourney makes more companies rely on AI-generated content for their posts, advertisements, and campaigns. Nevertheless, reliance on AI may cause a lack of originality and creativity in content creation and, therefore, make it repetitious. It is important to keep the balance between AI-generated and human-generated content in order to maintain the brand's identity.
- **Misinformation and Deepfakes:** Advanced AI technologies allow producing highly believable fakes in the form of images, video, and audio (known as deepfakes). The problem of spreading misinformation through AIbased content poses significant risks for both users and brands in terms of damaging reputations and misleading audiences. Therefore, AI detection should be considered an important investment for brands and companies.
- **AI Explainability and Black Boxes:** The modern AI technologies applied by various companies including Instagram and TikTok work as "black boxes". In other words, companies may not have full knowledge about how algorithms operate, resulting in the inability to optimize their marketing campaigns.
- **Content Saturation and Competitiveness:** The rise of AI has simplified content production, hence making the process much faster than it ever was before. This has led to

an oversupply of content every day. With this level of saturation, brands find it hard to differentiate themselves and grab the attention of their audiences. Creativity is the key to staying ahead of the competition.

- **AI Biases:** AI learns from data and past experiences. In case of any kind of bias in the data, this bias will be reflected in the AI algorithm. AI marketing campaigns might target some specific audience and overlook others. This can lead to bias, discrimination, and brand image problems.
- **Loss of Human Element in Communication:** The AI-based solutions used for marketing are automated and rely on algorithms that may not have emotions. This means that there could be a loss of human element in communication. Over-reliance on AI could lead to a robotic communication strategy, which could hurt the relationship between the brand and users.
- **Poor Data Quality and Accuracy:** AI-based marketing strategies are highly dependent on the data used to train these technologies. Poor data could lead to inaccurate results from these technologies.

This could impact how effective the AI technology is in making marketing decisions.

- **Difficult to Measure Impact of AI on Performance:** While AI technology helps improve the efficiency of marketing activities, measuring the effectiveness of AI technology can be challenging. It might be difficult for businesses to attribute changes in user engagement and sales directly to AI technology.
- **Advanced AI Application Cost:** Though entry-level AI systems are easily available, advanced AI applications entail substantial expenditure on technology, hardware, and manpower. This can pose a serious threat to smaller companies looking to compete against their bigger rivals.

2. **Basics of Artificial Intelligence**

Artificial Intelligence is an area that has many smaller parts and each part helps make smart systems. The main parts are Machine Learning, Deep Learning, Natural Language Processing and Robotics.

2.1 Machine Learning

Machine Learning is a part of Artificial Intelligence that makes machines learn from data and get better at what they do over time. This means they do not need to be programmed for every task. Machine Learning models look for patterns in data. Make decisions based on those patterns.

There are three kinds of Machine Learning:

- **Supervised Learning:** The model learns from data that has labels.
- **Unsupervised Learning:** The model looks for patterns in data that does not have labels.
- **Reinforcement Learning:** The model learns by trying things and getting rewards or penalties.

Machine Learning is used a lot on media for:

- Recommending content to users
- Finding spam and fake accounts
- Understanding how users behave
- Showing ads to the people

Machine Learning gets better and better as it gets more data, which makes it really good for things like social media that are always changing.

2.2 Deep Learning

Deep Learning is an advanced kind of Machine Learning that uses special networks with many layers. These networks can handle a lot of data. Find complex patterns.

Deep Learning is really good at:

- Recognizing images and videos
- Understanding speech
- Understanding language

For example Deep Learning helps automatically tag images on social media and makes voicebased interactions better. It also helps find hurtful content.

2.3 Natural Language Processing

Natural Language Processing helps machines understand and use human language. This makes Artificial Intelligence systems more interactive and user-friendly.

New Natural Language Processing systems use models that can understand context give human-like answers and do complex language tasks.

Natural Language Processing is used on media for:

- Chatbots and virtual assistants
- Understanding how users feel about things
- Translating languages
- Making content

Natural Language Processing helps platforms understand what users think and make their services better.

2.4 Robotics

Robotics is about making machines that can do tasks using sensors and motors. Even though Robotics is not directly related to media it helps with automation and collecting data that supports Artificial Intelligence systems.

Robotics is used a lot in industries like healthcare, manufacturing and logistics. It helps make Artificial Intelligence systems work better.

3. New Developments in Artificial Intelligence

Artificial Intelligence has changed a lot in the few years and new technologies have had a big impact on social media.

3.1 Generative Artificial Intelligence

Generative Artificial Intelligence is one of the important new developments in Artificial Intelligence. It lets machines make original content like text, images, audio and videos.

Some popular Generative Artificial Intelligence tools are:

- **ChatGPT:** A conversational chatbot powered by artificial intelligence that provides natural language responses and is capable of performing various tasks such as writing, programming, and communicating.
- **DALL-E:** An AI algorithm that can create pictures using texts, allowing users to create visual content effortlessly.
- **Midjourney:** AI-based software that uses text prompts to create impressive artwork, used extensively for designing and creating.

These tools are used on media for:

- Writing captions and posts
- Making images and graphics
- Making marketing content

Artificial Intelligence and Machine Learning are used to make these tools better and better. Artificial Intelligence is really good, at making content and it is used a lot on social media. Machine Learning helps make sure the content is good and relevant.

3.2 Recommendation Systems

Recommendation systems are artificial intelligence (AI)-based algorithms that leverage user data and behavior patterns to make recommendations. Such systems consider users' actions such as likes, shares, comments, and viewing time to understand their preferences.

For instance:

- Instagram recommends videos according to users' preferences
- Facebook orders news feed stories
- TikTok utilizes AI technology in its "For You" page

Such recommendation systems enhance user interaction with relevant content.

3.3 Computer Vision

Computer vision refers to technology that allows computers to recognize and interpret visual information from the real world in forms of pictures and videos. This technology is applied on social media platforms to achieve tasks including:

- Facial recognition
- Detection of objects
- Augmented Reality (AR) filters

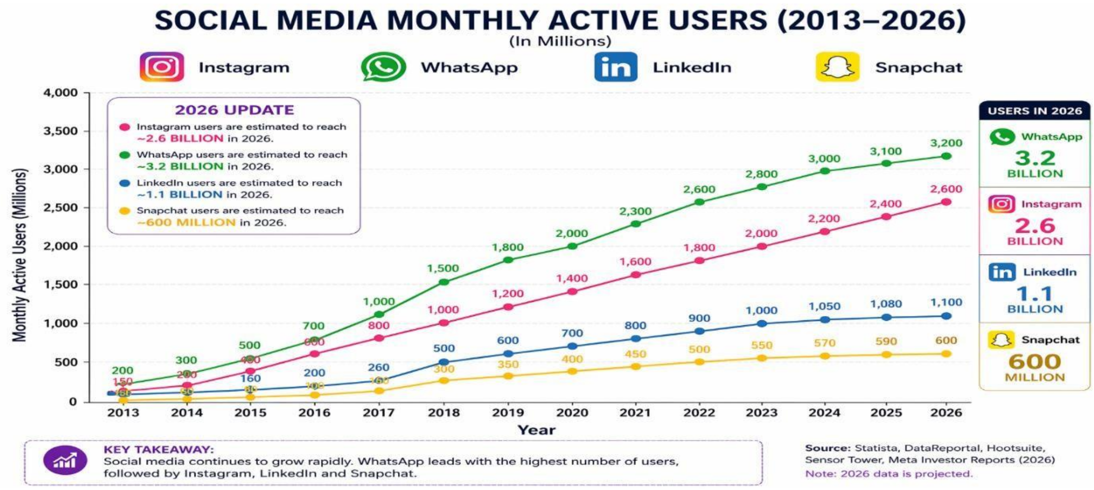
Such applications include filters in Snapchat and Instagram, automatic tagging of photos, and content moderation.

3.4 Artificial Intelligence in Social Media

Social media platforms rely heavily on Artificial Intelligence. AI has progressed from basic automation to intelligent decision-making systems that can learn from user behavior, generate content, and automate tasks. Popular social media platforms like Facebook, Instagram, TikTok, LinkedIn, and Pinterest have incorporated AI to better cater to their users' needs. Earlier, AI could only perform simple tasks like recommendations and automation. Now, thanks to recent developments, AI has become a lot smarter, and social media platforms are benefiting from it. Facebook

The AI algorithms of Facebook help in running all aspects of this popular social media platform, including recommendation of content, face recognition, suggesting friends, and sending highly-targeted ads. Facebook also uses AI to remove any fake account and detect spam and harmful contents.

- **Instagram:** Instagram uses AI to analyze images, and videos, and suggests content based on the interest of the users. Its Explore tab and Reels feature work based on AI. AI analyzes user actions like likes, comments, and watch time to give relevant content suggestions. AI is also used to enhance filters, edit images, and moderate content.
- **Snapchat:** The platform takes advantage of AI in the form of computer vision, which detects the face of users and adds instant filters using AR technologies. The smart filters enable more user engagement and offer interactivity. AI is also used for tracking user activities and suggesting content.
- **LinkedIn:** LinkedIn employs AI technology for recommending professional connections, jobs, and curating user feeds. Using AI algorithms, LinkedIn recommends job opportunities and connects users based on their user profiles, skillset, and activity. AI is also used by recruiters to find ideal candidates.
- **Pinterest:** Pinterest is one of the most famous AI-based applications. Pinterest recommends content to users according to their interests and actions. Using Pinterest Lens, people can look for products by uploading an image. A significant majority of Pinterest users base their purchases on recommendations generated through AI.



4. Literature Review

In years the stuff written about artificial intelligence and human-made content on social media has changed a lot. This change is because technology is advancing quickly and people are using media differently. Artificial intelligence-generated content has gotten really good due to improvements in learning, natural language processing and certain models like transformer-based architectures. These technologies let machines look at a lot of user data find patterns and make content that sounds like it was made by a human and even has a creative and emotional tone. Many studies show that AI tools can now make captions, posts, images and even videos with little help from humans. This is changing how content is traditionally made. Researchers stress that this change has made things more efficient lowered costs and allowed for content in real-time on a big scale. Intelligence and human-generated content are both important. At the time content made by humans is still really important for keeping things genuine, trustworthy and emotionally connected on social media. Research suggests that while artificial intelligence is great at being fast and handling amounts of data human creators are better at making original content that is aware of the context and culture.

Recent studies compare how much engagement content gets and show that users often think human-made content is more trustworthy and relatable. This is especially true, in areas that need experience storytelling or making ethical judgments. However, models that combine humans and AI to make content are becoming popular. They combine the best of both to make productivity and creativity better.

Artificial intelligence-generated content and human-generated content are both useful. Recent studies are looking at how AIGC affects the way people use things and how platforms work. These systems use intelligence to pick what people see which gets them to use the platform more but it also causes problems like people only seeing what they want to see and not getting all the facts. There are also worries about people using automated content to trick others or send spam. People who study this are talking about the need for platforms to be honest about what's real and what is not and to make sure the things that are made by artificial intelligence are fair.

To deal with this platform are making rules about what's made by artificial intelligence and what is not. The new information we have says that AIGC and the things people make do not have to compete with each other. They can work together. What we need to do is make sure we use artificial intelligence in a good way and that people and artificial intelligence work well together. We need to make rules that help us get things done quickly and make sure everything is real and honest so we can have a system for digital content that is fair and good, for everyone.

4.1 Applications of AI in Social Media

The use of social media has expanded far beyond a mere medium of communication. The platforms have now taken up several business tasks ranging from marketing and customer support to e-commerce. AI (Artificial Intelligence) plays a vital role in the same through the power of automation and analysis.

Some significant applications of AI in social media are described below:

- **Social Media Advertising:** AI has revolutionized advertisement and marketing on the social platforms by taking care of several crucial processes. Earlier, the advertisements were designed and placed manually with the help of minimal information such as age, gender, and locations. Today, however, AI can automate the entire procedure by generating ads, analyzing the user activity, and even suggesting the best performing ads. AI also makes predictions regarding the ads' future success and optimizes the campaign accordingly.

- **Marketing:** Artificial Intelligence has made digital marketing a whole new concept altogether, where the entire process has become easier through automation. From creating content to improving customer engagement, marketers today have access to powerful tools which can design and post the posts automatically using AI. AI is transforming digital marketing into a smarter and more automated process. Marketers can now use AI to understand customer preferences, create content, and improve engagement. AI tools can generate posts, captions, and even images using technologies like DALL·E and Midjourney.
- Marketing has evolved into AI-driven or Generative Marketing, where AI not only analyzes data but also creates complete marketing strategies and content automatically.
- **Social Insights:** AI-powered tools provide deep insights into user behavior, preferences, and trends. Earlier, businesses could only analyze past data, but now AI can predict future trends and user actions. It helps companies understand what type of content will perform best and how to improve their strategies.
- AI now offers predictive analytics, real-time performance tracking, and competitor analysis, helping businesses make faster and smarter decisions.
- **Security and Justice:** AI plays an important role in maintaining security on social media platforms. It helps detect spam, fake accounts, and fraudulent activities. Earlier systems were limited to basic filtering, but modern AI systems are much more advanced.
- AI can now detect fake news, deepfake videos, and harmful content like hate speech in real time, ensuring a safer and more reliable platform.
- **Automation:** Automation is one of the biggest advantages of AI in social media. Tasks such as content scheduling, posting, replying to messages, and managing accounts can be automated using AI tools. Automation has become intelligent automation, where AI can decide the best time to post, generate content automatically, and respond to users based on context.
- **Social Listening:** Social listening helps businesses monitor what people are saying about their brand on social media. Earlier, it was limited to tracking keywords, but AI has improved this process significantly. AI can now understand user emotions and sentiment (positive, negative, neutral) from comments and posts. This helps businesses improve customer satisfaction and brand reputation.
- **Chatbots:** Chatbots are widely used for customer interaction on social media. Earlier chatbots were rule-based and could only provide fixed responses. Modern chatbots powered by models like GPT-4 can understand context, provide human-like responses, and offer personalized support, improving customer experience.
- **Computer Vision & AR:** AI employs computer vision to interpret images and video footage. Complex capabilities such as face filters, augmented reality (AR), and object recognition have become prevalent, particularly on social media platforms such as Snapchat and Instagram.

In addition to these, new applications of AI have emerged in recent years, such as AI-based content creation, influencer marketing, recommendation systems, and computer vision technologies. AI can now generate posts, images, and videos automatically, suggest trending topics, and provide personalized content feeds based on user behavior. Features like augmented reality filters and facial recognition further enhance user interaction and engagement.

4.2 How are AI Tools used

AI tools have really improved a lot in the few years especially with the new updates from 2024 to 2026. They can do a lot more than simple tasks now. AI tools are like helpers and partners that can be creative too. These days AI tools can make things work with different types of media like text, images, videos and audio and even change things in real-time to fit each persons needs. This makes them really useful and easy to use in different areas.

There are some AI tools that have been updated recently like ChatGPT, Google Gemini, Midjourney and Runway ML. We can use these AI tools in practical ways, such as:

- **Content Creation:** AI tools can make social media posts, captions, blogs and even complete documents for projects. They can also change the way they write based on the tone we want who we are writing for and what platform we are using.
- **Image and Video Generation:** AI tools can make good images and videos from just a few words. This is really helpful for people who make things for media, marketing and presentations and we do not need to be experts in editing.
- **Coding and Development:** AI tools can help us write, fix and make our code better. Whether we are working with things like Laravel or making machine learning projects AI tools can save us a lot of time.
- **Personalized Learning:** AI tools can explain things to us in a way that is easy to understand, like deep learning how people feel about things and how to write research papers. New AI models can even change how they explain things based on how we understand them.
- **Data Analysis and Research:** AI tools can look at a lot of data. Find important information. We can use them to review what other people have written see what is trending and do research.
- **Productivity:** AI tools can do tasks for us like writing emails making reports and managing our schedules. This can help us be more productive.

To use AI tools well we should:

Write specific instructions like "Explain how to do a sentiment analysis project step by step with code"

Check and improve what the AI tool gives us instead of just using it without thinking

Use AI tools together to get better results like combining text, images and videos

Use AI tools as helpers not as a replacement for ourselves AI tools like ChatGPT and Google Gemini can be really useful, in this way.

4.3. Benefits of Artificial Intelligence and Human-Generated Content on Social Media

Intelligence and human creativity work together to make social media platforms what they are today. Artificial intelligence helps with automation and making decisions based on data while human-generated content makes sure people can trust and connect with what they see. The following sections explain the benefits of using intelligence and human-generated content on social media.

4.3.1. Benefits of Artificial Intelligence-Generated Content on Social Media

Artificial intelligence-generated content, which uses tools like ChatGPT and Google Gemini has changed the way content is made managed and optimized on media platforms.

- **Getting More People Involved:** Intelligence looks at how users behave what they like and how they interact with content in real time. Then it makes content that people will like. This way people are more likely to like, share and talk about the content. Now artificial intelligence can even predict what will be popular so brands can stay relevant.
- **Saving Time and Getting More Done:** Artificial intelligence does tasks that take a lot of time like scheduling content, writing captions making hashtags and tracking how well content does. This means marketers can focus on things that need creativity and strategy.
- **Making Ads Smarter and Better:** Artificial intelligence helps brands reach the people by looking at who they are and how they behave. It makes sure the right content gets to the people at the right time. Artificial intelligence also makes ads work better in time which means more people do what the ads want them to do.
- **Making Content Right for Each Person:** Artificial intelligence can make content that is just right for different groups of people. This includes recommending things people might like making feeds that change and making ads that're just for them. This makes people like using media more.
- **Saving Money and Getting a Better Return:** Artificial intelligence reduces the need for big teams to make and manage content. It saves money. Helps businesses get more for what they spend.
- **Artificial Intelligence-Powered Chatbots and Customer Support:** Artificial intelligence chatbots answer peoples questions away which makes people happy. They are available all the time. Can talk to many people at once. This makes people like using media more.

4.3.2. Benefits of Human-Generated Content on Social Media

Even though artificial intelligence is getting better human-generated content is still important for making social media feel real and personal.

- **Being Real and Trustworthy:** Content made by humans is more real and trustworthy because it comes from experiences and feelings. People like content that feels real and personal.
- **Connecting with People on a Deeper Level:** Humans can make people laugh feel empathy and tell stories. This makes people connect with the content more which is hard for artificial intelligence to do.

- **Being Creative and Original:** Humans can think of ideas and make content that is unique. Unlike intelligence, which uses what it already knows humans can be creative.
- **Understanding Culture and Context:** Humans understand what is appropriate and what people like better than intelligence. This helps make content that's right for different groups of people.
- **Handling Topics:** Human-generated content is better for topics that need people to think critically or feel emotions. This includes issues, personal stories and taking care of a brands reputation.
- **Building a Personal Brand. Being an Influencer:** Influencers and content creators use human-generated content to build a loyal audience. They use their personality and authenticity to make people like them.
- **Being Flexible and Adaptable:** Humans can change content based on what people say what is popular or what happens unexpectedly. They can change the tone, style and message easily than artificial intelligence.
- **Trusting Humans to Make Decisions:** People often trust what other humans say, review and recommend when making decisions. This makes human-generated content very valuable, for marketing and making brands credible.

5. Methodology:

This research is about looking at content made by intelligence and people on social media platforms. Artificial intelligence has gotten really good at making things and using many kinds of media so the way we do research has changed. We used to look at text. Now we use complicated methods that involve a lot of information and models. This study uses research methods with intelligence tools to make sure our results are accurate and useful in real life. We are talking about intelligence and human-generated content here.

5.1 Data Collection

Getting data is very important for this research. The quality of our results depends on the quality of the data we collect. We get our data from media platforms like Instagram, LinkedIn and YouTube.

How We Collect Data

We pick posts by hand based on whats popular or what people are talking about. This way our data shows what people are really interested in. We do this to make sure our data on intelligence and human-generated content is relevant. We also use automated tools to collect a lot of data. This includes things like captions, comments, likes, shares and timestamps. Using automation means we do not have to do work. We can get a lot more data, which makes our analysis of artificial intelligence and human-generated content better.

We have also made some changes to how we collect data:

- We now include images, videos and audio in our data not text to get a better understanding of artificial intelligence and human-generated content.
- We use intelligence tools to filter out data thats not relevant or is spam so our data on intelligence and human-generated content is clean.

- We can now collect data in time which means we can look at what is happening right now and understand how people are behaving on social media platforms with artificial intelligence and human-generated content.

This is important because we want to make sure our results are not biased and really show how people behave on media with intelligence and human-generated content.

5.2 Data Type and Classification

After we collect our data, we need to put it into groups that make sense. Artificial intelligencegenerated content is content made using intelligence tools.

Some key things about intelligence-generated content are:

- It uses language that is structured and grammatically correct which is different from humangenerated content.
- It often has patterns or phrases that repeat, which can be a sign of intelligence-generated content.
- It does not have stories or real-life context unlike human-generated content.
- It is often made to get a lot of engagement with keywords and hashtags which's a strategy used by artificial intelligence-generated content creators**.**

Human-generated content is content made by people. It is different from intelligence-generated content.

Some key things about human-generated content are:

- It has tone and personal opinions which's something artificial intelligence-generated content often lacks.
- It uses slang, informal language and cultural references which are common in humangenerated content.
- It has real-life experiences and stories which make human-generated content more relatable.

We have also made some changes to how we classify our data:

- We use machine learning to figure out if content's artificial intelligence or human-generated which helps us analyze artificial intelligence and human-generated content more accurately.
- We use models like BERT that understand context better than models so we can better understand artificial intelligence and human-generated content.
- We use a combination of labeling and artificial intelligence classification to get results, which is important for our study of artificial intelligence and human-generated content.

5.3 Tools and Technologies

We use tools to help us analyze our data. We use the programming language Python because it is powerful and flexible which is useful for analyzing intelligence and human-generated content.

We use libraries like Pandas to handle our data, NumPy for operations, SpaCy and NLTK for natural language processing and Matplotlib for data visualization all of which help us understand intelligence and human-generated content better.

We use machine learning and artificial intelligence models like Logistic Regression, Naive Bayes and LSTM which're useful for analyzing artificial intelligence and human-generated content. We also use models like Transformer Models that understand context, which's important for our study of artificial intelligence and human-generated content.

We have also made some changes to our tools

- We use -trained artificial intelligence models so we do not have to train them from scratch, which saves us time and effort in our analysis of artificial intelligence and human-generated content.
- We use cloud platforms to handle datasets, which's helpful for managing large amounts of data on artificial intelligence and human-generated content.
- We use AutoML tools to pick the models, which helps us find the effective way to analyze artificial intelligence and human-generated content.

5.4 Data Preprocessing

The data we get from media is often messy and noisy so we need to clean it up. The steps we take are:

- We remove things like URLs, emojis, special characters and symbols that're not relevant to our study of intelligence and human-generated content.
- We convert all the text to lowercase. It is consistent which makes it easier to analyze intelligence and human-generated content.
- We break the text into words or tokens which helps us understand the structure of intelligence and human-generated content.
- We remove words like "s" "the,"and". " Which are common in both artificial intelligence and human-generated content but do not add much value to our analysis.

We have also made some changes to our preprocessing:

- We convert emojis into sentiment values, which gives us insight into how people react to intelligence and human-generated content.
- We understand words based on the meaning of the sentence, which's important for accurately analyzing artificial intelligence and human-generated content.
- We can handle languages using intelligence tools, which helps us analyze intelligence and human-generated content from different languages.

With advancements in artificial intelligence sentiment analysis has become more advanced than just simple classification. Modern techniques provide insights into user emotions and the context of their feedback.

5.5 Sentiment Analysis

With advancements in artificial intelligence sentiment analysis has become more advanced than just simple classification. Modern techniques provide insights into user emotions and the context of their feedback.

1. Emotion Detection Emotion detection goes beyond positive, negative and neutral categories. It identifies human emotions like, happiness, anger, sadness, surprise and sarcasm

For example, a comment like "Wow thiss just great…" may seem positive but could actually be sarcastic. Advanced AI models can detect subtle emotional expressions making the analysis more accurate with emotion detection.

2. Context-Aware Models Traditional sentiment analysis relied on keyword matching, which often led to interpretations. Modern models use context- techniques. This means they understand the meaning of words based on the sentence.

For instance,

“This post is sick" could mean negative or positive depending on context.

Context-aware models interpret this correctly as positive if used in slang with context- models.

These models are based on architectures like transformer models. They consider the sentence rather than individual words for context-aware models.

3. Aspect-Based Sentiment Analysis Aspect-based sentiment analysis focuses on components of content. It does not analyze the text as one unit.

For example:

- Caption sentiment → Positive
- Comments sentiment → Mixed or negative

This helps in understanding which part of the content is performing well. It also shows which part needs improvement with aspect-based sentiment analysis. It is especially useful in media.

Users react differently to captions, visuals and comments on media.

- How Sentiment Analysis Works (Step-by-Step Process)
- The sentiment analysis process follows a pipeline

Step 1: Data Input

Text data such as comments, captions and replies are collected from social media platforms for data input.

Step 2: Preprocessing

The text is cleaned by:

- Removing URLs, emojis and special characters
- Converting text to lowercase * Tokenizing words

Step 3: Feature Extraction

Important features are extracted from the text. These features include keywords, phrases or embeddings that represent the meaning of the text for feature extraction.

Step 4: Sentiment Classification

A machine learning or AI model is used to classify the text into sentiment categories. These categories include negative, neutral or emotions with sentiment classification.

Step 5: Sentiment Scoring

Each text is assigned a sentiment score.

For example:

- Positive →1
- Neutral → 0
- Negative → -1

Advanced models may assign scores within a range. This range could be -1 to +1 for precision, with sentiment scoring.

Step 6: Aggregation

All sentiment scores are. Averaged. This determines the sentiment of:

- AI-generated content
- Human-generated content

6. Future Scope of AI in Social Marketing

The future of technology in media marketing companies is really exciting. Social media marketing companies will be able to create posts, images and videos more easily with technology. This means social media marketing companies will not have to put in much manual effort and their content will stay consistent.

There are many things technology can do for media marketing companies. One important thing technology can do for social media marketing companies is advanced customer targeting. Technology studies what people do online what people like and what people prefer. Then social media marketing companies can show the content to the right people at the right time. This makes people interact more with media marketing companies and it also makes people more likely to buy products or services from social media marketing companies.

Technology is also very useful for analysis for social media marketing companies. Social media marketing companies can understand what people will like in the future and what people will prefer. Social media marketing companies can even guess if their marketing campaigns will work. This helps social media marketing companies make decisions and feel more confident about their decisions.

Another big advantage of technology for social media marketing companies is the use of chatbots. Chatbots can talk to people anytime, day or night. Answer their questions right away. This makes people happier with media marketing companies and it saves time for social media marketing companies.

Technology also helps social media marketing companies improve their advertising campaigns in time. Technology constantly checks how the campaigns are doing. It can make quick changes like changing the budget or targeting the right people. This leads to results for social media marketing companies without much delay.

Personalized marketing is another benefit of technology for social media marketing companies. Social media marketing companies can show people content based on what they like making the experience more relevant and interesting for people. People are more likely to respond when they see content that matches their interests.

Technology also helps media marketing companies with influencer marketing. Social media marketing companies can find influencers by looking at their audience and how engaged they are. Technology can even find followers helping social media marketing companies make better decisions, about who to work with.

In addition, technology helps social media marketing companies understand what people think through feedback analysis. By looking at comments and reactions social media marketing companies can know what people like or dislike. This allows social media marketing companies to improve their products and services.

7. Findings

- **Hyper-Personalization for Each Person:** This is really cool because it creates an experience for every single person. It does this by looking at what they do what they like and what they are doing now. Then it gives them content and suggestions that are just right for them.
- **Helps in Making Smart Decisions:** This thing is really smart. It can tell what is going to happen in the future. It also gives ideas for content tells us when to post and helps plan things of time. This way we do not have to guess. We get better results.
- **Automating Customer Interactions:** We have things like chatbots and virtual assistants that answer customer questions away. They are available all the time so people can talk to them whenever they want. This makes talking to customers better.
- **Real-Time Campaign Optimization:** We always watch how our campaigns are doing. Then we make changes, to who we're talking to how much money we are spending and what we are saying. This helps us get results and not waste money.
- **Growth of Automated Content Creation:** There are tools that can make captions, images and videos fast. This makes it easier to create content. It looks the same on all our platforms.
- **Improving Influencer Marketing:** This helps us find the influencers. It checks if people are really looking at what they post and if they have followers. This makes our campaigns work better. We can trust them more.

Conclusion

This research shows that both Artificial Intelligence generated content and human generated content are important for media. Artificial Intelligence is good at doing things it can do a lot of work and it is consistent. It can also make decisions based on data, which's very useful. Artificial Intelligence generated content can make a lot of content look at what usersre doing and make marketing plans better. This makes Artificial Intelligence very useful for businesses that want to reach people get more people engaged and do better overall. But human generated content is still very important because it is creative it has feeling it is real. It can make real connections with people who use social media. Human content is better at telling stories understanding what is happening in a culture and being kind to people, which's important for keeping people trusting us and having long term relationships with them. From this study we can see that using Artificial Intelligence or just human content is not enough to get the best results. The best way to do things is to use a mix of both, where Artificial Intelligence helps and makes human creativity better. Artificial Intelligence can do tasks look at data and make things better while humans can focus on being creative making plans and connecting with people on an emotional level. This research also talks about problems like concerns issues with keeping data private and the risk of relying too much on machines. We need to solve these problems for Artificial Intelligence to keep growing in a way on social media. To keep growing over time companies need to focus on AI use being open and trusting users. They must also help their employees learn skills to keep up with changing technologies. This way companies can make things and still be successful for a long time on social media. Companies should be fair and responsible when using AI and changing with technologies. This helps build trust with users and keeps companies doing well. They must make sure they are using AI in a way that's good, for everyone. In conclusion the future of media content is about Artificial Intelligence and humans working together. Companies that can combine the things, about Artificial Intelligence with the creative ideas of humans will be more likely to succeed in the world of digital media.

References:

1. Aldous, K., Salminen, J., Farooq, A., Jung, S. G., & Jansen, B. (2026). Understanding user engagement with cross-platform social media content. *ACM Transactions on the Web*. Advance online publication. https://doi.org/10.1145/3756014
2. Galiano-Coronil, A., Aguirre Montero, A., López Sánchez, J. A., & Díaz Ortega, R. (2024). Exploring social responsibility, social marketing and happiness using artificial intelligence, automated text analysis and correspondence analysis. *Management Decision, 62*(2), 549–574. https://doi.org/10.1108/MD-01-2023-0099
3. Hua, Y., Niu, S., Cai, J., & Althoff, T. (2024). Generative AI in user-generated content. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems* (Article 471, pp. 1–7). Association for Computing Machinery. https://doi.org/10.1145/3613905.3636315
4. Israel, M. J., & Amer, A. (2022). Rethinking data infrastructure and its ethical implications in the face of automated digital content generation. *AI and Ethics, 2*(4), 687–698. https://doi.org/10.1007/s43681-022-00169-1
5. Møller, A. G., Romero, D. M., Jurgens, D., & Aiello, L. M. (2026). The impact of generative AI on social media: An experimental study. *Scientific Reports, 16*, Article 40110. https://doi.org/10.1038/s41598-026-40110-8
6. Park, J., Oh, C., & Kim, H. Y. (2024). AI vs. human-generated content and accounts on Instagram: User preferences, evaluations, and ethical considerations. *Technology in Society, 79*, Article 102705. https://doi.org/10.1016/j.techsoc.2024.102705
7. Silva, R. T., Santos, M. A., & Ferreira, L. C. (2026). A linguistic comparison between human- and AI-generated content. *Journal of Language and Social Psychology, 45*(1), 112–129. https://doi.org/10.1177/0261927X251296908
8. Smith, A. J., & Jones, B. L. (2026). Is it cake or is it AI? A systematic review of human uncertainty in distinguishing generative artificial intelligence content. *Computers in Human Behavior, 174*, Article 108437. https://doi.org/10.1016/j.chb.2026.108437
9. Wang, X., & Chen, Y. (2026). Visible sources and invisible risks: Exploring the impact of AI disclosure on perceived credibility of AI-generated content. *Journal of Science Communication, 25*(1), Article A09. https://doi.org/10.22323/2.25010209

10. Dhavaleshwar, S. B. (2024). Algorithmic consumer dynamics: The role of artificial intelligence in personalizing social media marketing strategies for enhanced customer experience in corporate India. *Indian Journal of Marketing Research, 14*(3), 45–59.
11. Gupta, S., & Pandey, M. (2021). Deconstructing consumer engagement patterns: A comparative framework of human vs. automated content interfaces on Indian retail handles. *Journal of Retailing and Consumer Services, 60*, Article 102451. https://doi.org/10.1016/j.jretconser.2021.102451
12. Jain, R., & Aggarwal, S. (2020). AI-driven brand building vs. organic human-generated content: Evaluating metrics for social platforms in emerging economies. *Indian Journal of Commerce & Management Studies, 11*(2), 18–27. https://doi.org/10.18843/ijcms/v11i2/03
13. Kumar, V., Rajan, B., & Venkatesan, R. (2024). Resolving the privacy-personalization paradox: Consumer trust development in AI-enhanced social media spaces. *International Journal of Information Management, 75*, Article 102731. https://doi.org/10.1016/j.ijinfomgt.2023.102731
14. Pathan, M., Jarad, R., Bhavsar, D., Yadav, M., Suryavanshi, P., & Moholkar, N. (2024). Artificial intelligence in social media marketing. *International Journal of Research Publication and Reviews, 5*(12), 2780–2784. http://www.ijrpr.com
15. Siau, K. L., & Yang, Y. (2020). Impact of artificial intelligence, robotics and machine learning on sales and marketing. *Journal of Database Management, 31*(3), 45–63. https://doi.org/10.4018/JDM.2020070104

REAL-TIME MONITORING AND PROCTORING IN A LAN-BASED ONLINE EXAMINATION SYSTEM

S. Abishek, V. Srinivasan and N. Senthil Kumar
Department of Informatics (Data Science),
Periyar Maniammai Institute of Science and Technology (Deemed to be University),
Thanjavur – 614206, India
Corresponding author E-mail: abisachi002@gmail.com, srinivasan95615@gmail.com, profnsk22@gmail.com

Abstract

A campus examination is a strange workload: for fifty minutes nothing much happens, and then within a window of seconds two hundred browsers all try to write their final answer at once. Off-the-shelf quiz platforms struggle here because they are designed for the public web and tuned for steady traffic rather than thunder-clap bursts. This paper describes a quiz system built the other way round, for a single shared local area network. Two design choices do most of the heavy lifting. First, the per-student question shuffle is a pure function of the registration number and the test ID, generated on demand through an xmur3 hash feeding a mulberry32 PRNG, so the database does not have to remember any permutations. Second, SQLite runs in Write-Ahead Logging (WAL) mode, which lets readers and writers stop blocking each other when the timer expires. On top of that sits a proctoring component that listens to focus and visibility events on the DOM and reports them in real time, plus the surrounding scaffolding—registration, scheduled tests, attempt counting, and recovery after a browser crash. Across our load tests the per-student database footprint dropped by roughly 98%, reconnection settled below 85 ms, and the system held together at concurrencies where the cloud-style baseline did not.

Keywords: Decentralized Assessment, Deterministic Ran-domization, Real-Time Analytics, WebSockets, Write Ahead Log-ging, Behavioral Proctoring.

Introduction

Online quizzes are now timetabled and graded, but they typically run on the same web stack that powers shopping sites—REST endpoints behind a CDN, a relational database somewhere far away, the open internet between every student and the server [1]. On the public web that is fine; inside a university lab it is fragile. Every browser share one uplink, the round-trip to the cloud crosses equipment the institution does not control, and a five-second hiccup loses real student work. The literature acknowledges this in passing but rarely treats the local network as a first-class deployment target.

Recent designs have replaced repeated polling with Web-Socket push [2], which is well understood. The database side has had less attention: cloud-native systems hide write contention by spreading load across replicas, but on a single-machine deployment with one SQLite file, the closing minute of an exam is where contention shows up and there is nowhere for it to fan out to. The shuffling story is similar—the conventional way to stop students copying from a neighbour is to hand each one a stored permutation, which grows linearly in the cohort size even though it could be regenerated from a seed.

This work targets the LAN case directly: server, students, and invigilator all on the same switch, no inter-net assumed during the exam. The contributions are: (i) storage-free per-student randomisation through a deterministic xmur3→mulberry32→Fisher–Yates pipeline; (ii) submission-storm tolerance via SQLite WAL mode; (iii) live proctoring that surfaces focus loss, copy/paste, and tab switches as they happen [3]; (iv) lifecycle-aware scheduling with attempt limits, server-enforced timer expiry, and crash-resilient reconnection; and (v) two delivery modes side by side—synchronous live tests and self-paced scheduled tests.

Literature Review and Research Gaps

Clark *et al.* [5] measured cloud-hosted Learning Management Systems under synchronous load and reported that the latency curve bends sharply around 150 concurrent users—not a slow drift but a knee. Pimentel and colleagues [6] looked at long-polling and found that cohort size and server load come out roughly quadratic. On the storage side, Hipp's documentation [7] is candid that in rollback-journal mode a writer takes an exclusive lock on the entire database file, so a few hundred submissions in the same instant serialise into a queue. Smith *et al.* [8] quantified the storage cost of the standard "shuffle and save" approach: for 100 questions and 500 students, permutation data alone reaches roughly 45 MB, an $O(N \cdot Q)$ overhead that exists only because the server is asked to remember a sequence it could regenerate from a seed.

Table 1 pairs each gap with the corresponding choice in this work.

Table 1: Critical literature survey and research gap identification

Concept	Limitation	Proposed Solutions
Cloud LMS via REST	External internet dependency, latency prone.	Local zero latency Web Socket framework.
HTTP Long Polling	Server thread bottlenecks under concurrency.	Event-driven Socket.IO
Standard DB locks	Locks during simultaneous submissions.	SQLite Write Ahead Log.
Server-side shuffling	Database bloat, $O(N \cdot Q)$overhead.	Deterministic PRNG shuffling.
Static delivery	No lifecycle / attempt tracking.	Scheduled orchestration with timer enforcement.

System Architecture and Topology

The system is organised into three layers, with boundaries drawn for the LAN case. The browser layer is a React single-page application served through Vite; it paints the UI, watches focus and visibility events, and keeps client state in React Context. The server layer is one Node.js process: Express handles the small amount of REST traffic (mostly authentication) while Socket.IO carries the bidirectional WebSocket conversation that the rest of system flows through. Live session state lives in memory; anything that has to survive a restart goes to the persistence layer. The database layer is a single SQLite file in WAL mode, holding accounts, test definitions, answers, violation records, and attempt rows.

Socket.IO is configured to prefer the websocket transport with polling as a fallback only; once the WebSocket handshake clears, per-message HTTP framing disappears. The buffer cap is raised to 10 MB so that a question bank with embedded base64 images is delivered in one piece. The wire protocol uses approximately twenty-eight named events, split across client-to-server (joining, answering, cheat reporting, focus and navigation updates, scheduled-test joining, submit and auto-submit, and administrative commands), server-to-client (question display, test start/end, ban/unban, lock/unlock, broadcast, waiting count, cheat warnings, fullscreen enforcement, and scheduled-test events), and administrative state-sync events feeding the dashboard, throttled to one snapshot every 40 ms so that a synchronised burst does not turn the UI into a strobe light.

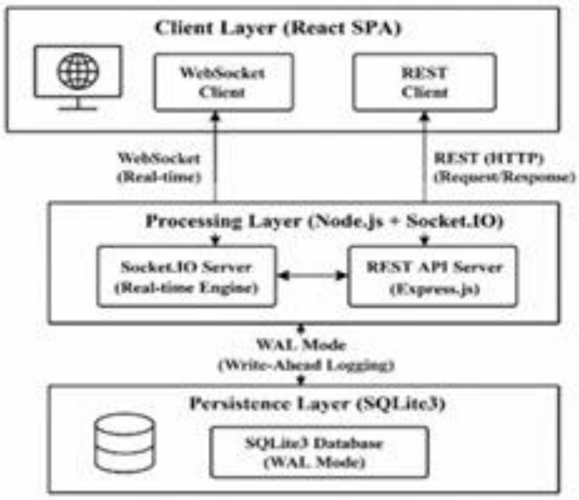


Figure 1: Three-Tier Architecture Overview

Concurrency and Persistance

Write contention here is not evenly distributed. It concentrates at the start and end of the test; in SQLite's default rollback-journal mode the closing seconds form a queue in which readers block writers and writers block readers. The escape hatch is three PRAGMAs at startup: journal_mode =WAL, synchronous = NORMAL, and foreign_keys = ON. WAL changes the shape of concurrency entirely [4]—new transactions append to a separate WAL file rather than mutating the main file in place, so a reader (the admin dashboard polling for the leaderboard, say) always sees the last committed snapshot and is never held back by a writer. synchronous = NORMAL relaxes fsync discipline just enough to make group commits fast while keeping the WAL durable across normal process exits.

Answers go through an INSERT ...ON CONFLICT (reg, test_id, qid) DO UPDATE rather than a plain insert. The triple is unique, so a flaky network that retransmits an answer, a student who navigates back and changes a choice, or a reconnect that replays a queued payload all collapse to a single row—no application-level deduplication step. Six tables carry the load: students, student_sessions, student_answers, violations, scheduled_tests, test_attempts, plus a small config bag. Migrations are handled by ALTER TABLE ...ADD COLUMN statements wrapped in try/catch blocks so that an older database file opens cleanly under a newer build.

Deterministic PRNG and State-Bloat Elimination

Shuffling is the obvious anti-cheating measure: randomise the question order per student and the choice positions per question. The textbook implementation stores both shuffles, an $O(N \cdot Q)$ cost. The approach taken here skips the storage entirely: each shuffle is a pure function of the student's identity and the test ID, so it is regenerated whenever needed and per-user database cost falls to $O(1)$.

The composite string reg | TEST_ID is converted into a 32-bit seed using a MurmurHash3-style finalizer (xmur3), which exhibits the avalanche property the rest of the design depends on—two registration numbers differing by a single digit produce uncorrelated seeds. The seed feeds mulberry32, a small Weyl-sequence generator that steps its internal state by the constant 0x6D2B79F5 and applies an xorshift multiply mixer. Using a fully deterministic PRNG (rather than Math.random) is what makes crash recovery cheap: the same seed reproduces the same stream forever.

The PRNG drives a backwards Fisher–Yates pass at two independent levels. The question order is seeded with reg | testId so each (student, test) pair gets its own ordering; the choice order is seeded with reg | qid so even when two students hit the same question the option layout is independent. Splitting the seeds keeps the shuffles uncorrelated, and because both seeds are derived from values that already exist (a registration number, a question ID), nothing has to be written. Compactly,

S(r, t) = FisherYates(Q, mulberry32(xmur3(r t))) (1) where S(r, t) is the sequence delivered to student r on test t.

Authentication, Sessions and Recovery

Account passwords are stored as bcrypt hashes at cost factor 10, and session tokens are 256 random bits from crypto. randomBytes(32). Inputs are validated at the boundary: registration numbers must be all digits, e-mail addresses must parse, and passwords must clear a minimum length. A successful login writes a token into student_sessions and returns it to the browser; refreshing the page calls /student/verify to trade the token for a logged-in session without prompting again. The administrator side is intentionally simpler—a shared password sent through an x-admin-pass header, with success setting an httpOnly cookie.

LANs drop connections; the reconnect path is layered. The client re-emits its identity on the new socket, the server restores the in-memory user record, pulls existing answers out of student_answers, regenerates the question order from the original seed in O(1), and drops the student into the next unanswered question with no visible interruption.

Scheduled Test Orchestration

Two delivery modes are supported because campus exams really do come in two shapes. In live, administrator-driven mode, the administrator picks a question bank, configures parameters, and pushes a start signal; every student in the waiting room begins at the same instant. In scheduled, student-paced mode, a test is configured with a window (starts_at, ends_at), a duration cap, an attempt limit, and an optional assignment filter. Students see those tests on their dashboard and start them inside the window with their own countdown. Every (student, test) pair walks through eligibility, timer calculation, attempt creation, the active session, and completion. Eligibility checks the time window, the assignment list (per-student, per-department, or per-range), and remaining at-tempts. The countdown is min(duration, window remaining), so a student who joins late never overruns the close. Completion happens in three ways: the student submits, answers the last question, or a 5-second server-side poll notices the deadline has passed and auto-submits on their behalf. All checks are evaluated on the server, so editing the request payload in the browser does not gain access.

Multi- Tiered Proctoring

Most exam software treats proctoring as forensic: collect logs now, look at them later. The approach here is the inverse—the client watches behaviour live and pushes events the instant they happen, with enforcement attached. Detection runs at the DOM level and splits into two tiers. Tier

1 (counted strikes) catches deliberate-looking breaches: a tab/visibility switch via the Document Visibility API, a window blur for Alt-Tab and external focus theft, and clipboard misuse (copy, paste, cut, PrintScreen) intercepted on the key-board event path. The system clipboard is wiped programmatically with navigator.clipboard.writeText('') so a copied snippet cannot survive the violation. Tier 2 (silent suppression) drops common escape keystrokes via e.preventDefault() without notifying the student: F1–F12, Alt-Tab, Ctrl-Tab, the right-click context menu, DevTools shortcuts, and browser navigation shortcuts.

The strike threshold is configurable; the default is four. Strikes 1–3 fire a cheat:warning showing the count and remaining headroom. From strike four the client flips into lockedByCheat: questions are CSS-blurred so they cannot be read, and the student stays stuck until an administrator clears the lock. An administrator can also remove a student outright via admin:ban. Every violation is written to the violations table with a timestamp and type, which is what makes appeal reconstruction possible. The fullscreen feature uses the standard browser Fullscreen API; the administrator broadcasts fullscreen:enforce, every connected client requests fullscreen, and any subsequent exit is logged as a violation. Mobile Safari and a handful of older devices that do not implement the API are treated as already fullscreen rather than refused entry.

The administrative dashboard is a single page split into eight panels: an overview of live counters; a waiting room with per-student approve/remove controls; test control with start/stop/force-advance and the fullscreen and free-navigation toggles; a participants table with online status, current question index, score, focus state, and violation count; a bank manager with import/export and CRUD on reusable question banks; a question writer with image support and per-question time overrides; scheduled tests with the assignment editor; and a leaderboard that updates through the same WebSocket as everything else.

Implementation and Performance

The implementation uses Node.js (ES Modules) with Ex-press 4.x and Socket.IO 4.7 on the server; React 18.3 with React Router 6.26, Framer Motion 11.3 and Tailwind CSS 3.4 on the client; SQLite3 in WAL mode through the sqlite/sqlite3 packages; bcryptjs and Node crypto for security; Vite 5.3 for the build; and ExcelJS for XLSX export. The bank manager handles bulk CSV/JSON import, an in-app editor that supports two to six choices per question with base64 images and per-question time overrides, persistent reusable banks, bulk export, save-test-as-bank, and test duplication. Anything coming in is normalised: a question is rejected if its ID is missing, the text is blank, there are fewer than two choices, no choice is marked correct, or any field has the wrong type. Two navigation modes are available—linear (forward only) and free (forward and backward, preserving previously chosen answers)—and as with every other rule, navigation is enforced server-side.

Crash recovery: A browser crash mid-exam should not cost the student their progress, and with the deterministic PRNG it does not. The new session reconnects; the server pulls already-stored answers out of the WAL-backed database, reconstructs the answer set, re-runs the registration ID through xmur3 to confirm the question order matches, and drops the student back at the next unanswered question. End-to-end this completes in under 85 ms on our test hardware; a cloud-based design that has to make external API calls for the same operation typically lands closer to 1200 ms.

Comparative analysis: The proposed system was compared against a traditional cloud-API setup and a stock SQLite (rollback-journal) configuration with 200 simulated concurrent students on a single machine (Table II). The headline line is the database footprint—0.8 MB against 45.2 MB, a 98.2% drop—because there are simply no permutation rows to write. WAL mode brings the worst-case write latency from 850 ms down to 12 ms during the closing-second submission rush. Per-user memory falls from 92 KB (a stored shuffle plus state) to 1.2 KB.

Table 2: Comparative performance (200 concurrent users)

Metric	Cloud	Std. SQLite	Proposed
State recovery latency	1200 ms	450 ms	85 ms
DB payload overhead	45.2 MB	45.2 MB	0.8 MB
Concurrent write delay	350 ms	850 ms	12 ms
Memory per user	92 KB	92 KB	1.2 KB
Max stable concurrency	80	120	500+

Table 3 shows the scaling behaviour. P99 latency stays under 100 ms even at 500 concurrent students and the curve is essentially linear, which is what an $O(1)$ per-user PRNG cost predicts.

Table 3: Scalability under increasing load

Metric	50	100	200	500
Avg. response (ms)	3	5	12	28
P99 latency (ms)	8	15	35	72
CPU (%)	4	8	18	42
Memory (MB)	45	52	68	112

Security and Resuls Export

Every user-supplied field is validated server-side: digit-only registration numbers, RFC-style e-mail addresses, minimum-length passwords hashed with bcrypt at cost 10, and structural checks on question payloads. Every incoming answer is checked against the question the server believes the student is currently on; if the question ID does not match the submission is dropped silently. Deadlines are enforced server-side: the moment Date.now() > u.timerEnd, the at-tempt is auto-submitted regardless of what the client thinks. Sessions are bound to a unique cryptographic token, and a fresh login wipes earlier sessions for the same registration through deleteSessionsByReg, which means a stolen token expires the next time the genuine user logs in. The administrator cookie is set with httpOnly and sameSite: lax, closing the obvious CSRF vector.

After the exam, results are exported as XLSX through ExcelJS (per-test, with student details, scores, and attempt counts), per-question accuracy is sorted lowest-first to flag questions that need rewriting, the violations table is filterable by student or by test for academic-integrity review, and a live leaderboard updates through the WebSocket while the test is still running.

Conclusion and Future Work

This paper described the design and implementation of a real-time, LAN-resident assessment platform. The two ideas doing most of the work are (i) treating per-student randomisation as a deterministic function rather than a stored array, and (ii) running SQLite in WAL mode so submission storms stop being write-contention events. Together they collapse the per-user database footprint, make crash recovery fast enough to be invisible, and remove the failure modes that come with depending on the public internet inside a campus lab. The two delivery modes—live invigilator-driven and student-paced scheduled—between them cover the cases university examinations actually produce, and the proctoring layer turns observation into enforcement at the moment the violation happens rather than a week later.

Future directions worth pursuing include webcam-based proctoring with on-device face presence and gaze-direction inference; adaptive testing layered on Item Response Theory; multi-node clusters with Redis pub/sub for institutions that need to spread load across machines; an offline-first PWA so a momentary network outage does not interrupt the exam at all; and tamper-proof certification through a distributed ledger for independently verifiable result claims.

References

1. We Create Problems (WECP). (2026). *How proctoring platforms detect and prevent cheating?*
2. SQLite Consortium. (2026). *Write-ahead logging*. https://www.sqlite.org/wal.html
3. Patel, V. (2024). Scalable cloud architectures for intelligent mobile learning platforms. *Journal of Software Development and Educational Informatics*.
4. VideoSDK. (2024). *Long polling vs WebSockets: A developer's guide to real-time solutions*.
5. Hipp, D. R. (2024). *SQLite database architecture: Journaling modes and concurrency control*. SQLite Official Documentation.
6. Pimentel, R., Santos, L., & Oliveira, K. (2022). Real-time state synchronization via HTTP long polling: Performance analysis under concurrent load. *Web Engineering Review, 18*(2), 112–128.
7. Clark, A., Davis, R., & Thompson, M. (2023). Cloud-based assessment frameworks and latency constraints in synchronous examinations. *Journal of Educational Technology Systems, 51*(3), 287–305.
8. Smith, J., Anderson, P., & Williams, T. (2023). Memory overhead analysis in server-side shuffled arrays for large-scale academic assessments. *ACM Transactions on Computing Education, 23*(4), 1–18.

A COMPREHENSIVE REVIEW OF APPLIED RESEARCH IN SCIENCE, ENGINEERING, AND TECHNOLOGY

Pooja Narang[1] and Vandana Grover[2]

[1]Department of Botany, School of Sciences, IFTM University, Moradabad 244102(U.P), India

[2]Department of Chemistry, BBD NIIT, Lucknow 226028, (U.P) India

Corresponding author E-mail: pooja.narang@iftmuniversity.ac.in, vandanagrover26@gmail.com

Abstract

Applied research has a special and vital place in the broader area of science. Unlike pure or basic research, which is largely driven by intellectual curiosity and the quest of theoretical understanding, applied research primarily focuses on solving practical problems and providing knowledge that can be quickly used in real-world settings. An extensive examination of applied research in the domains of science, technology, and engineering is provided in this chapter. It delineates the methodological frameworks that set applied inquiry apart from other types of research, examines its philosophical and epistemological foundations, and charts its historical development from the Industrial Revolution to the current era of digital transformation and artificial intelligence. Particular attention is paid to interdisciplinary collaboration, applied research ethics, study design strategies, and the methods by which findings are transformed into benefits for society. Case studies from the domains of biology, environmental engineering, materials science, and computers are analysed to illuminate the practical elements of the topic. An outlook on future opportunities and difficulties in applied research is presented at the end of the chapter, emphasizing the significance of this field for both human progress and technological civilization.

Keywords: Applied Research, Scientific Methodology, Engineering Design, Technology Transfer, Interdisciplinary Inquiry, Research Ethics, Innovation, Translational Science.

1. Introduction

Research is civilization's cornerstone. From the first systematic observations of the natural world to the sophisticated computational modeling of laboratories in the twenty-first century, human desire to understand, predict, and control reality has produced knowledge of astounding range and depth. Because it bridges the gap between abstract knowledge and real-world application, applied research plays a unique role in this massive undertaking by transforming basic scientific discoveries into products that revolutionize public policy, commercial operations, and daily life.

Although helpful for analysis, the distinction between fundamental and applied research has always been somewhat artificial. Some of the most innovative scientific projects are motivated by both curiosity and practical applications, as demonstrated by Pasteur's famous quadrant model (Stokes, 1997). Nonetheless, the applied research tradition maintains a unique identity that is defined by an emphasis on problem-solving, applicability to particular contexts, stakeholder involvement, and a commitment to generating measurable outcomes. In particular, such tradition is examined in this chapter.

Despite being frequently combined under the general acronym STEM, science, technology, and engineering are separate but intricately linked epistemic communities. Engineering uses scientific and mathematical concepts to design and construct answers to specific problems; science produces propositional information about the world; and technology creates objects and systems that increase human capacity. These communities are connected by applied research which facilitates the exchange of information, resources, and skills across academic boundaries.

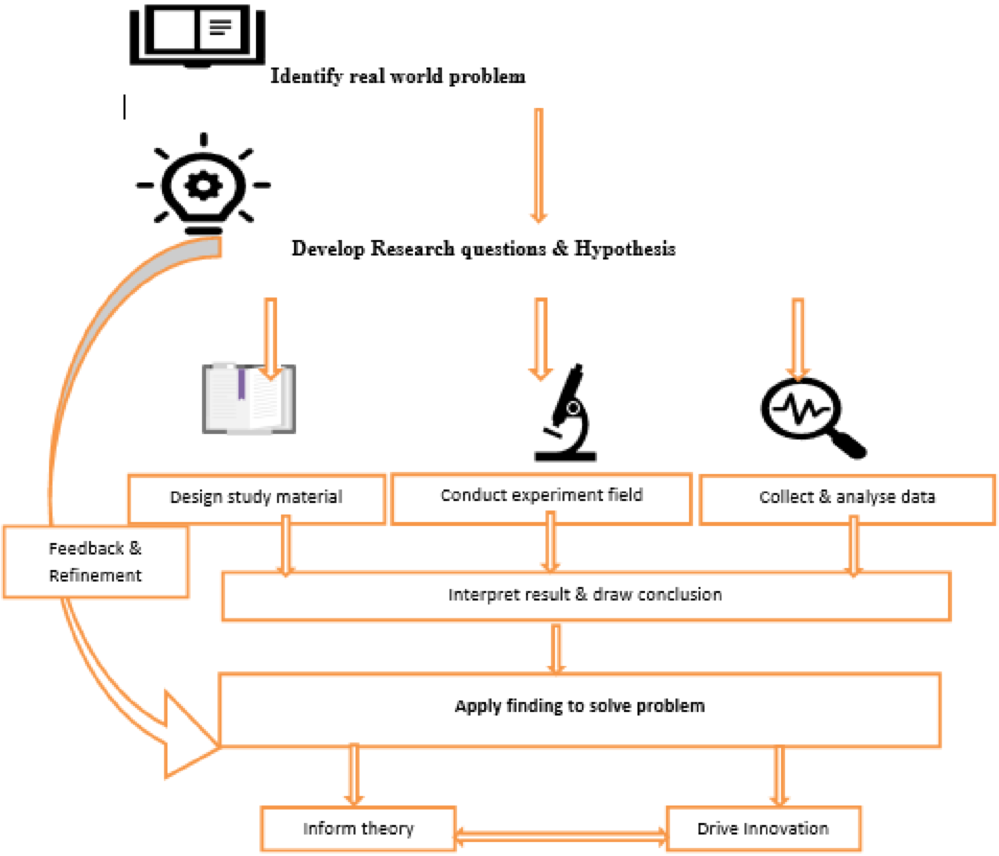


Figure 1: The Applied research Process

2. Historical and Conceptual Overview of Applied Research

2.1 History and Initial Development

Applied research has a long history. Archimedes' study of buoyancy, which was motivated by the practical challenge of recognizing contaminated gold, finest exemplifies the applied tradition's marriage of theoretical curiosity and practical necessity (Dijksterhuis, 1987). Mesopotamia's agricultural achievements and ancient Rome's hydraulic engineering are examples of systematic, if pre-scientific, attempts to apply natural knowledge for human benefit. Modern applied research as an independent undertaking is usually credited to the Industrial Revolution of the eighteenth and nineteenth centuries. For example, the development of the steam engine was a result of both mechanical intuition and systematic experimentation.

The meticulous study of steam properties conducted by James Watt and chemist Joseph Black is a prime example of the productive interplay between scientific understanding and engineering design that would later define applied research (Cardwell, 1971).

2.2 Twentieth-Century Institutionalization

In the twentieth century, applied research was formally institutionalized through the establishment of industrial laboratories, government research bodies, and university-industry partnerships. Bell Laboratories was established in 1925 and became a prominent hub of applied science, producing inventions ranging from information theory to the transistor while keeping a close eye on the practical needs of the telecom industry (Gertner, 2012).Similarly, the establishment of national laboratories during and after World War II, including Oak Ridge, Argonne, and the UK's National Physical Laboratory, demonstrated the government's recognition of applied research as an essential national resource. In the postwar decades, Vannevar Bush developed his well-known "linear model" of innovation, which holds that basic research drives applied research, which in turn leads to technological improvement and commercial application (Bush, 1945).

This paradigm established a robust institutional framework for research finance and organization that continues to impact scientific systems worldwide, despite substantial adjustments and criticism in the decades that followed.

2.3 Modern Conceptualizations

Modern ideas of applied research are more dynamic and complicated than the linear model implies. Research that is context-driven, transdisciplinary, and reflexively accountable to multiple stakeholders is referred to as mode 2 knowledge generation (Gibbons *et al.,* 1994). This idea holds that in addition to receiving input from basic science, applied research actively determines research aims, generates new theoretical insights, and participates in the co-production of knowledge with practitioners, policymakers, and communities.

Applied inquiry is evolving, as seen by the emergence of concepts like challenge-led research, open science, and responsible research and innovation (RRI). These frameworks emphasize the necessity of anticipating societal consequences, ensuring broad access to research outputs, and aligning scientific endeavours with societal values and priorities (Owen *et al.,* 2012).

3. Foundations of Philosophy and Epistemology

3.1 Commitments to Ontology and Epistemology

The foundation of applied research is a broadly realist ontology, which holds that there is an external reality with properties that can be known by methodical investigation, at least in theory. This dedication to realism underscores the idea that research findings will have predictive power and practical utility, setting applied science apart from solely constructivist or interpretive methods. Applied research uses a pluralist toolset in terms of epistemology. Particularly in the physical and biological sciences, positivist presumptions that knowledge is based on observable, quantifiable events and that causal linkages may be proven through controlled experimentation remain significant.

However, qualitative, interpretive, and interactive approaches that defy reduction to strictly quantitative methodologies are often incorporated into applied research in engineering, environmental science, and health (Cresswell, 2014).

3.2 The Relevance Issue

The dilemma of relevance—how to make sure that research topics, methods, and outputs are truly responsive to practical requirements rather than being motivated by the internal logic of academic disciplines—is a unique epistemological challenge for applied research. There are institutional and cognitive aspects to this issue. Cognitively, researchers trained in disciplinary traditions might not have the contextual knowledge needed to recognize the most urgent real-world issues or comprehend the limitations that must be met by proposed remedies. Institutionally, academic reward structures that privilege publication in high-impact journals may create incentives misaligned with the production of practically useful knowledge (Gibbons *et al.,* 1994).

Various strategies have been developed to address the relevance problem, including participatory research design, embedded researchers, knowledge brokers, and boundary organisations that span the interface between research and practice (Cash *et al.,* 2003).

3.3 Theory and Practice in Applied Inquiry

Applied research is not atheoretical. On the contrary, robust theoretical frameworks are essential for guiding the design of applied studies, interpreting their results, and generalising findings beyond the specific contexts in which they were generated.

In applied research, the interaction between theory and practice is best understood as dialectical and iterative: theoretical models guide practical interventions, which in turn produce empirical data that may support, contradict, or change theoretical claims. Design science research in engineering and information systems, which alternates between building artifacts and developing theoretical knowledge about design principles, is a good example of this iterative dynamic (Hevner *et al.,* 2004). Similar to this, action research in the social and organizational sciences generates both theoretical understanding and practical improvements by combining cyclical processes of planning, action, observation, and reflection.

4. Applied Research Methodological Frameworks

4.1 Quantitative Approaches

Applied research in the natural sciences, engineering, and medicine continues to rely heavily on quantitative methodologies. A solid foundation for determining causal linkages and optimizing interventions is provided by experimental design, which includes response surface techniques, factorial experiments, and randomized controlled trials. Researchers can find significant patterns in complicated, high-dimensional data sets using statistical methods like machine learning and analysis of variance. Finite element analysis, computational fluid dynamics, signal processing, and systems modelling are examples of quantitative techniques used in engineering settings. These tools speed up the design cycle and eliminate the need for expensive physical prototypes by allowing researchers to replicate physical events with high fidelity (Bathe, 2006).

4.2 Mixed and Qualitative Approaches

In applied research, qualitative approaches are becoming more widely acknowledged, especially in fields where human behavior, social context, and subjective experience are important for the creation and uptake of solutions. Focus groups, semi-structured interviews, ethnographic observation, and case study analysis offer significant contextual information that cannot be obtained through quantitative methods alone. Complex practical problems are especially well-suited for mixed methods research, which methodically integrates quantitative and qualitative approaches within a single study or program. Depending on the research question and context, sequential designs in which qualitative investigation influences the creation of quantitative instruments and concurrent designs in which quantitative and qualitative data are gathered concurrently and combined in analysis offer different benefits (Cresswell, 2014).

4.3 Methods Based on Computation and Data

Data-driven inquiry is a new paradigm in applied research that has emerged as a result of the growth of digital data and improvements in computing capacity. Researchers can find patterns, make predictions, and uncover connections in data sets of a size and complexity that would be unmanageable with traditional analytical procedures thanks to machine learning, deep learning, and artificial intelligence tools. Computational methods have revolutionized applied research in a variety of domains, including precision agriculture, materials design, drug development, and climate modeling. Significant advances in computationally assisted applied research include high throughput screening, genome-wide association studies, and protein folding simulation utilizing programs like AlphaFold (Jumper *et al.,* 2021).

These advancements create significant issues regarding the interpretability and reliability of machine learning models in high-stakes applications, as well as the connection between data-driven pattern identification and mechanistic knowledge.

Table 1: Comparison of Key Methodological Approaches in Applied Research

Method / Approach	Primary Application & Characteristics
Randomised Controlled Trial	Gold standard for causal inference in biomedical and behavioural research; minimises confounding through random allocation.
Finite Element Analysis	Numerical simulation of physical systems (structural, thermal, fluid); essential in mechanical and civil engineering design.
Ethnographic Observation	Deep contextual understanding of user behaviour and social dynamics; informs human-centred design and technology adoption.
Mixed Methods Design	Integrates quantitative and qualitative data to address complex applied problems requiring both breadth and depth.
Machine Learning / AI	Pattern recognition and predictive modelling in large, high-dimensional data sets; increasingly central to data-driven applied research.
Action Research	Cyclical inquiry combining intervention and reflection; used in organisational, educational, and community-based applied research.

5. Interdisciplinary Collaboration in Applied Research

5.1 The Imperative of Interdisciplinarity

Complex realworld problems rarely respect disciplinary boundaries. Climate change, pandemic preparedness, sustainable urban development, and the governance of artificial intelligence each demand the integration of knowledge, methods, and perspectives from multiple scientific and professional traditions. Applied research is, by its nature, disposed towards interdisciplinary collaboration, and the most significant advances in the field have frequently emerged from productive encounters between disciplines.

For instance, molecular biology, computer science, robotics, statistics, and ethics would not have been able to work together to complete the Human Genome Project (Collins *et al.,* 2003). Similar to this, decades of interdisciplinary research in immunology, biochemistry, materials science, and clinical medicine were used in the creation of mRNA vaccine technology against COVID-19. When properly funded and institutionally supported, transdisciplinary applied research can be creative, as demonstrated by these examples.

For instance, molecular biology, computer science, robotics, statistics, and ethics would not have been able to work together to complete the Human Genome Project (Collins *et al.,* 2003). Similar to this, decades of interdisciplinary research in immunology, biochemistry, materials science, and clinical medicine were used in the creation of mRNA vaccine technology against COVID-19. These instances show how interdisciplinary applied research, when properly funded and institutionally supported, may be productive.

5.2 Difficulties with Multidisciplinary Work

Despite its potential, interdisciplinary research has unique organizational and epistemological difficulties. Collaborators may find it difficult to communicate and understand one another due to differences in discipline terminology, methodological conventions, and epistemic standards. Innovative research that deviates from accepted genre norms may be disadvantaged when interdisciplinary outputs are evaluated by discipline peer reviewers. Early-career researchers may be deterred from pursuing multidisciplinary pathways by departmental allegiances and career hierarchies.

Deliberate investment in team procedures, common conceptual frameworks, and institutional frameworks that facilitate boundary-crossing work is necessary for successful multidisciplinary applied research. Clear goal-setting, explicit role-and-responsibility negotiation, time set aside for relationship-building, and leadership that values diversity of perspective are all important components of successful interdisciplinary collaboration, according to the team science literature (Stokols *et al.,* 2008).

5.3 Industry-University Collaborations

One important organizational route for applied research is university-industry cooperation. These collaborations make use of university research groups' intellectual resources while granting access to infrastructure, market knowledge, and industrial experience. They have played a key role in the growth of knowledge-intensive businesses including digital communications, biotechnology, and semiconductor manufacturing. Partnerships between universities and business, however, also bring up significant issues regarding the management of intellectual property, the protection of academic freedom, the alignment of interests, and the possibility that commercial priorities will skew research agendas. Realizing the advantages of these collaborations while controlling their hazards requires effective governance structures (Perkmann *et al.,* 2013).

6. Applied Research's Ethical Considerations

6.1 Applied Researchers' Social Responsibility

Applied researchers are moral actors whose choices have important social ramifications; they are not just technical specialists. The creation of dual-use technologies, such as genetic editing, autonomous weaponry, and surveillance systems, highlights the moral significance of applied research decisions. Deep concerns about justice, autonomy, and human dignity are raised by the application of artificial intelligence in healthcare, the use of big data analytics in criminal justice, and the deployment of new agricultural biotechnologies in food systems.

Respect for people, beneficence, and justice are the cornerstones of the research ethics tradition, which is based on seminal texts like the Declaration of Helsinki and the Belmont Report (1979). These guidelines, which were first created in the context of biomedical research involving human beings, have been expanded upon and modified to handle the unique ethical difficulties associated with engineering, computational, and environmental research.

6.2 Ethical Innovation and Research

Developed as part of European science policy, the Responsible Research and Innovation (RRI) framework offers a thorough method for integrating ethical thinking into applied research practice. In addition to inclusive stakeholder participation, reflexivity regarding researchers' own presumptions and values, and responsiveness to public concerns and preferences, RRI demands anticipatory governance and the prospective assessment of potential societal repercussions (Owen *et al.,* 2012). Ethics review committees, impact assessment instruments, public engagement procedures, and emerging technology regulatory frameworks are just a few of the ways that RRI principles have been put into practice. The application of RRI concepts to quickly developing practical research frontiers is both necessary and challenging, as demonstrated by the governance of CRISPR-Cas9 gene editing (Jasanoff *et al.,* 2021).

6.3 Quality Control and Research Integrity

The societal value of applied research is based on its integrity. Plagiarism, fabrication, and falsification compromise the validity of the body of information that practical applications rely on. More subtle challenges to research integrity, such as selective reporting, p-hacking, poor disclosure of conflicts of interest, and insufficient transparency about procedures and data, have garnered increasing attention in recent years in addition to overt misbehaviour (Ioannidis, 2005). Pre-registration of study ideas, sharing of data and code, and reporting of negative outcomes are examples of open science practices that are being promoted more and more as systemic solutions to the integrity issues that applied research faces.

7. Examples of Case Studies

7.1 The Evolution of CRISPR-Based Treatments

Initially described by Doudna and Charpentier (2012) as a bacterial immune mechanism, the CRISPR-Cas9 system has quickly evolved into a flexible platform for applicable research in biotechnology, agriculture, and medicine. CRISPR-based treatments were in clinical trials for diseases like sickle cell disease, beta-thalassaemia, and several types of cancer within ten years after their first characterization. The quick transition from basic discovery to applied investigation, the intense interdisciplinary collaboration spanning biochemistry, genetics, medicine, and bioethics, the complicated regulatory environment governing clinical translation, and the profound ethical questions raised by the possibility of heritable human genome editing are all characteristics of applied research in the life sciences that this case exemplifies.

The difficulties of responsible innovation in frontier applied research are exemplified by the wider social and ethical ramifications of CRISPR technology, such as concerns about fair access, augmentation versus therapy, and ecological dangers of gene drives in wild populations.

7.2 Water Purification and Environmental Engineering

An estimated 2.2 billion people do not currently have access to securely managed drinking water services, making access to clean water one of the major issues of the twenty-first century (WHO/UNICEF, 2021). A wide range of water purification technologies, from traditional filtration and chlorination to sophisticated oxidation processes, membrane filtration, and solar disinfection, have been produced by applied research in environmental engineering. The development of low-cost, community-scale water treatment systems for deployment in low-income contexts illustrates the importance of designing research for contextual appropriateness as well as technical effectiveness. Effective solutions must be affordable, robust under field conditions, maintainable with locally available skills and materials, and acceptable to the communities that adopt them. Research approaches that engage potential users in co-design and prioritise locally defined needs over technically elegant solutions have demonstrated superior outcomes in practice (Montgomery and Elimelech, 2007).

7.3 Artificial Intelligence in Medical Diagnostics

The application of machine learning to medical imaging represents one of the most significant domains of applied AI research. Deep learning algorithms have demonstrated diagnostic accuracy comparable to or exceeding that of specialist clinicians for conditions including diabetic retinopathy, dermatological malignancies, and pulmonary embolism detected on computed tomography scans (Rajpurkar *et al.,* 2022). Beyond algorithmic performance, a variety of practical research challenges have emerged as a result of the translation of these laboratory accomplishments into clinical practice. Serious issues about equality are raised by dataset bias, as models trained primarily on data from wealthy, predominately white populations perform poorly on underrepresented groups. Additional issues that have required consistent applied research effort to solve include model interpretability, regulatory approval, liability, and clinician-AI collaboration. This instance serves as an example of the idea that in applied research, technical performance is essential but insufficient to have a positive real-world influence.

7.4 Cutting-Edge Renewable Energy Materials

Applied research in materials science and engineering has increased dramatically as a result of the world's shift to renewable energy. A prime example of translational applied research is perovskite solar cells, which have shown power conversion efficiencies above 25% in lab settings and provide notable manufacturing cost savings over traditional silicon photovoltaics (NREL, 2023). Fuel cells, thermoelectric energy harvesting, and lithium-ion battery technology have all advanced as a result of applied materials research. The lengthy and non-linear process from materials discovery to commercial implementation is demonstrated by these examples. Stability, scalability, and device integration issues have necessitated consistent, iterative applied research. The use of machine learning and high-throughput simulation in computational materials discovery to screen potential materials is an example of how cutting-edge applied materials research integrates computational and experimental methods.

8. Future Directions and New Challenges

8.1 Convergence and Systems Approaches

The main direction of contemporary applied research is convergence, or the merging of once distinct scientific disciplines into coherent research and technological platforms. The convergence of nanotechnology, biotechnology, information technology, and cognitive science—known as NBIC convergence has the potential to transform human-machine interaction, manufacturing, communications, and medicine in ways that are difficult to anticipate (Roco and Bainbridge, 2003). Systems techniques, which examine complex events as emergent properties of interacting components rather than as the sum of independently understood elements, are becoming more and more crucial in navigating this converging landscape.

8.2 Public and Open Science

The open science movement, which encourages broad access to research data, methods, and publications, is transforming the institutional environment of applied research. Open data repositories, open access publication models, and open-source software platforms are lowering barriers to entry, accelerating the dissemination of knowledge, and enabling new types of distributed, collaborative inquiry. Citizen science broadens the scope of applied research into populations and contexts that are unreachable by professional researchers alone by involving non-professionals in data collecting, analysis, and interpretation.

8.3 Global Issues and Research Driven by Challenges

A shift in applied research funding and practice toward challenge-led inquiry is being driven by the scope and severity of global issues such as climate change, biodiversity loss, antimicrobial resistance, pandemic preparedness, and sustainable development. Programs like the European Green Deal research agenda and the Welcome Trust's Our Planet Our Health initiative serve as examples of this strategy, which emphasizes the direct relevance of research to specific societal concerns and encourages transdisciplinary collaboration across conventional boundaries.

Innovative approaches to research organization, funding, assessment, and governance are necessary for challenge-led applied research in order to handle the complexity, urgency, and value of the issues it tackles. One significant field of applied research in science policy and innovation studies is the creation of suitable institutional frameworks for challenge-led inquiry.

8.4 Artificial Intelligence as an Instrument and Subject of Useful Study

In modern applied research, artificial intelligence plays two roles: it is an object of applied research in and of itself and a potent instrument for accelerating discoveries across scientific domains. Computer scientists, social scientists, ethicists, and legal scholars must continue to conduct practical research on the safety, dependability, interpretability, justice, and governance of AI systems. AI is revolutionizing a number of applied research fields, including medicine development, materials design, and climate modelling. One of the key problems of applied research in the upcoming decades will be comprehending and handling the ramifications of AI-enabled applied research, including issues of responsibility, intellectual property, and the evolving nature of scientific competence.

Conclusion

In science, technology, and engineering, applied research is essentially a human and social endeavour rather than just a technical one. In addition to the characteristics of the natural world, it is influenced by institutional structures, power dynamics, values, and interests. The most important applied research is methodologically advanced and contextually anchored, technically inventive and ethically reflective, rigorous and pertinent at the same time. Using case studies from the fields of biomedicine, environmental engineering, artificial intelligence, and materials science, this chapter has explored the philosophical underpinnings, methodological frameworks, collaborative structures, historical development, and ethical aspects of applied research.

Genuine problem orientation, the importance of interdisciplinary collaboration, the necessity of ethical reflection and stakeholder engagement, the potential of computational approaches, and the crucial role of institutional frameworks that support the translation of knowledge into practice are among the common principles that emerge across these diverse domains. The relevance of applied research in science, technology, and engineering will only increase as the world's problems worsen in the twenty first century. In addition to technical innovation, overcoming these obstacles will call for institutional ingenuity, political resolve, and a strong dedication to the human values that applied research ultimately advances.

References

1. Bathe, K. J. (2006). *Finite element procedures*. Prentice Hall.
2. Bush, V. (1945). *Science: The endless frontier: A report to the President on a program for postwar scientific research*. United States Government Printing Office.
3. Cardwell, D. (1971). *From Watt to Clausius: The rise of thermodynamics in the early industrial age*. Heinemann.

4. Cash, D. W., Clark, W. C., Alcock, F., Dickson, N. M., Eckley, N., Guston, D. H., Jager, J., & Mitchell, R. B. (2003). Knowledge systems for sustainable development. *Proceedings of the National Academy of Sciences, 100*(14), 8086–8091. https://doi.org/10.1073/pnas.1231332100
5. Collins, F. S., Morgan, M., & Patrinos, A. (2003). The Human Genome Project: Lessons from large-scale biology. *Science, 300*(5617), 286–290. https://doi.org/10.1126/science.1084564
6. Cresswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.
7. Dijksterhuis, E. J. (1987). *Archimedes*. Princeton University Press.
8. Doudna, J. A., & Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science, 337*(6096), 816–821. https://doi.org/10.1126/science.1225829
9. Gertner, J. (2012). *The idea factory: Bell Labs and the great age of American innovation*. Penguin Press.
10. Gibbons, M., Limoges, C., Nowotny, H., Schwartzman, S., Scott, P., & Trow, M. (1994). *The new production of knowledge: The dynamics of science and research in contemporary societies*. SAGE Publications.
11. Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly, 28*(1), 75–105. https://doi.org/10.2307/25148625
12. Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLOS Medicine, 2*(8), e124. https://doi.org/10.1371/journal.pmed.0020124
13. Jasanoff, S., Hurlbut, J. B., & Saha, K. (2021). CRISPR democracy: Gene editing and the need for inclusive deliberation. *Issues in Science and Technology, 37*(2), 37–43.
14. Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., ... & Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature, 596*, 583–589. https://doi.org/10.1038/s41586-021-03819-2
15. Montgomery, M. A., & Elimelech, M. (2007). Water and sanitation in developing countries: Including health in the equation. *Environmental Science & Technology, 41*(1), 17–24. https://doi.org/10.1021/es0724856
16. National Renewable Energy Laboratory. (2023). *Best research-cell efficiency chart*. Retrieved from https://www.nrel.gov/pv/cell-efficiency.html
17. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont report*. U.S. Department of Health, Education, and Welfare.
18. Owen, R., Macnaghten, P., & Stilgoe, J. (2012). Responsible research and innovation: From science in society to science for society, with society. *Science and Public Policy, 39*(6), 751–760. https://doi.org/10.1093/scipol/scs093
19. Perkmann, M., Tartari, V., McKelvey, M., Autio, E., Brostrom, A., D'Este, P., ... & Sobrero, M. (2013). Academic engagement and commercialisation: A review of the

literature on university-industry relations. *Research Policy, 42*(2), 423–442. https://doi.org/10.1016/j.respol.2012.09.007

20. Rajpurkar, P., Chen, E., Banerjee, O., & Topol, E. J. (2022). AI in health and medicine. *Nature Medicine, 28*, 31–38. https://doi.org/10.1038/s41591-021-01614-0
21. Roco, M. C., & Bainbridge, W. S. (Eds.). (2003). *Converging technologies for improving human performance*. Kluwer Academic Publishers.
22. Stokes, D. E. (1997). *Pasteur's quadrant: Basic science and technological innovation*. Brookings Institution Press.
23. Stokols, D., Hall, K. L., Taylor, B. K., & Moser, R. P. (2008). The science of team science: Overview of the field and introduction to the supplement. *American Journal of Preventive Medicine, 35*(2), S77–S89. https://doi.org/10.1016/j.amepre.2008.05.002
24. WHO/UNICEF. (2021). *Progress on household drinking water, sanitation and hygiene 2000–2020: Five years into the SDGs*. World Health Organization and United Nations Children's Fund.

CADMIUM TELLURIDE THIN FILMS: SYNTHESIS, CHARACTERIZATION, AND FUNCTIONAL PROPERTIES FOR OPTOELECTRONIC AND ENERGY APPLICATIONS

Prashant A. Chate[1] and Dattatray J. Sathe[2]

[1]Department of Chemistry, J.S.M. College, Alibag (M.S.) India

[2]Department of Chemistry, KIT College of Engg (Autonomous), Kolhapur (M.S.) India.

Corresponding author E-mail: pachate04@rediffmail.com, djsathe77@gmail.com

1. Introduction

Cadmium telluride (CdTe) is one of the most prominent and widely utilized II–VI semiconductor materials, particularly in the domain of thin film optoelectronics and photovoltaic energy conversion. Over the past few decades, CdTe has established itself as a cornerstone material for thin film solar cells due to its near-ideal optoelectronic properties, ease of fabrication, and scalability for industrial production. The combination of a suitable direct band gap, high optical absorption coefficient, and favorable charge transport characteristics makes CdTe thin films highly efficient for converting solar energy into electrical energy [1-3].

CdTe crystallizes in a cubic zinc blende structure with a lattice constant of approximately 6.48 Å and a density ranging from 5.85 to 6.20 g/cm³. This crystal structure provides a well-ordered lattice that supports efficient charge carrier transport. The material exhibits a direct band gap of approximately 1.45eV at room temperature, which is very close to the theoretically optimal value for maximum solar energy conversion efficiency according to the Shockley–Queisser limit. This band gap allows CdTe to absorb a significant portion of the solar spectrum, particularly in the visible region. One of the most important optical characteristics of CdTe is its high absorption coefficient, typically greater than 10^5 cm^{-1} for photon energies above the band gap [4-5]. This means that CdTe can absorb about 90–99% of incident sunlight within a thickness of only 1–2 µm, which is significantly thinner compared to conventional crystalline silicon. As a result, CdTe thin films enable the fabrication of lightweight, flexible, and cost-effective solar cells with reduced material consumption. From an electrical perspective, CdTe exhibits excellent semiconducting properties. The intrinsic carrier concentration is relatively low, typically in the range of 10^{13} to 10^{14} cm^{-3}, but can be increased through doping. The carrier mobility is relatively high, with electron mobility ranging from 100 to 1100 $cm^2/V\cdot s$ and hole mobility between 40 and 100 $cm^2/V\cdot s$, depending on film quality and preparation conditions. CdTe can be engineered to exhibit both n-type and p-type conductivity, although stable p-type conductivity is more commonly utilized in device fabrication. The electrical resistivity of CdTe thin films generally lies in the range of 10^3 to 10^7 $\Omega\cdot cm$, which can be tailored through doping and annealing processes. The significance of CdTe thin films is most evident in their application in photovoltaic technology, where they have become one of the leading materials for thin film solar cells. CdTe-based solar cells have achieved laboratory efficiencies exceeding 22%, while commercial modules typically exhibit efficiencies in the range of 18–20%. These values are among the highest for thin film technologies and are comparable to many crystalline silicon-based devices. Additionally, CdTe solar cells have a lower manufacturing cost, reduced energy payback time, and simpler fabrication processes, making them economically viable for large-scale deployment [6-8]

CdTe thin films are also highly significant in radiation detection applications due to the high atomic numbers of cadmium and tellurium, which provide strong absorption of X-rays and gamma rays. This makes CdTe an excellent material for X-ray detectors, gamma-ray spectrometers, and medical imaging devices. Furthermore, CdTe is used in infrared detectors, photoconductive devices, and electro-optic modulators, owing to its favorable electronic structure and optical response. The thin film form of CdTe offers several advantages, including compatibility with a wide range of substrates such as glass, metal foils, and flexible polymers. Various deposition techniques can be employed for the synthesis of CdTe thin films, including close-spaced sublimation, chemical bath deposition, electrodeposition, sputtering, thermal evaporation, and vapor transport deposition. Among these, close-spaced sublimation and electrodeposition are widely used for large-scale industrial production due to their high deposition rates and uniform film formation. Typical film thickness ranges from 1 to 10 µm for solar cell applications, although thinner films are increasingly explored for advanced device architectures. Despite its numerous advantages, CdTe also presents certain challenges. The toxicity of cadmium and the limited availability of tellurium are important concerns for environmental sustainability and long-term scalability. However, modern CdTe technologies incorporate robust encapsulation methods and recycling strategies that minimize environmental impact. Additionally, ongoing research focuses on improving material utilization efficiency and exploring alternative compositions to address resource limitations. In recent years, significant advancements have been made in improving the performance of CdTe thin films through defect engineering, doping optimization, interface modification, and heterojunction design. For example, the incorporation of buffer layers such as CdS and the use of back contact materials like ZnTe or Cu-doped layers have significantly enhanced device efficiency and stability.

2. Synthesis of CdTe Thin Films by Chemical Bath Deposition (CBD)

CBD is a simple, low-cost, and scalable technique widely used for the preparation of semiconductor thin films, including CdTe. The method is particularly advantageous for large-area deposition, low-temperature processing, and uniform film growth [9]. In the case of CdTe, CBD involves the controlled release of cadmium (Cd^{2+}) and telluride (Te^{2-}) ions in an aqueous solution, leading to the formation and deposition of CdTe thin films on a suitable substrate through heterogeneous nucleation. In a typical CBD process for CdTe thin films, a cadmium salt such as cadmium chloride or cadmium acetate is used as the source of Cd^{2+} ions. The tellurium source is usually provided in the form of sodium tellurite or tellurium dioxide, which is reduced in situ to generate Te^{2-} ions. Since telluride ions are not stable in aqueous solutions, a reducing agent such as hydrazine hydrate, sodium borohydride, or ascorbic acid is used to convert Te^{4+} species into Te^{2-} ions.

To control the release of Cd^{2+} ions and prevent immediate precipitation, a complexing agent such as ammonia, triethanolamine, or ethylene diamine tetra acetic acid is added to the solution. These agents form stable complexes, thereby regulating the availability of free Cd^{2+} ions. The typical concentration of cadmium precursor is maintained in the range of 0.01–0.05 M, while the tellurium precursor concentration is usually kept between 0.005 and 0.02 M. The pH of the bath is adjusted to an alkaline range (pH 10–12) using ammonium hydroxide or sodium hydroxide, which facilitates the reduction of tellurium species and the formation of CdTe. Before deposition, substrates such as glass, indium tin oxide, or fluorine-doped tin oxide are thoroughly cleaned to ensure proper adhesion and uniform film growth. The cleaning process typically involves washing with detergent, rinsing with distilled water, ultrasonic cleaning in acetone and ethanol for about 10–15 minutes each, followed by drying in air or nitrogen. A clean substrate surface enhances nucleation and leads to better film quality.

The chemical bath is prepared by first dissolving the cadmium salt in distilled water, followed by the addition of the complexing agent under constant stirring. The tellurium precursor is then introduced into the solution along with the reducing agent to generate Te^{2-} ions. The solution is stirred continuously to ensure homogeneity. Once the bath is ready, the cleaned substrate is immersed vertically into the solution to allow uniform deposition on both sides. The deposition process is typically carried out at temperatures ranging from 50°C to 80°C for duration of 60–180 minutes. During this period, Cd^{2+} ions slowly released from the complex react with Te^{2-} ions generated in the solution to form CdTe, which deposits onto the substrate. The film thickness can be controlled by varying deposition time and precursor concentrations, and typically lies in the range of 200 nm to 2 μm. Initially, nucleation occurs at active sites on the substrate, followed by growth and coalescence of grains to form a continuous thin film. The growth mechanism in CBD involves both ion-by-ion deposition and cluster-by-cluster deposition. In the ion-by-ion mechanism, Cd^{2+} and Te^{2-} ions react directly on the substrate surface, resulting in smooth and uniform films. In the cluster-by-cluster mechanism, CdTe particles form in the bulk solution and subsequently deposit on the substrate, leading to rougher and more porous films. The dominance of either mechanism depends on bath conditions such as supersaturation, temperature, and ion concentration. The growth kinetics of CdTe thin films can be described by the relation: $d = kt^n$ where d is the film thickness, t is the deposition time, k is the growth rate constant, and n is the growth exponent. A value of n close to 1 indicates reaction-controlled growth, while n = 0.5

suggests diffusion-controlled growth. The growth rate is initially high due to the abundance of ions but decreases over time as reactants are consumed.

After deposition, the films are removed from the bath, rinsed with distilled water to eliminate loosely adhered particles, and dried. Post-deposition treatment is often necessary to improve film quality. Annealing is typically carried out at temperatures between 300°C and 450°C, often in the presence of $CdCl_2$, which enhances crystallinity, increases grain size, and reduces defects. This treatment significantly improves the optical and electrical properties of the films. Under optimized conditions, CdTe thin films prepared by CBD exhibit polycrystalline structure, good adhesion, uniform coverage, and grain sizes in the range of 100–500 nm. The films show strong optical absorption, suitable band gap, and improved electrical conductivity after annealing, making them highly suitable for photovoltaic and optoelectronic applications.

3. Crystallographic Properties of CdTe Thin Films

The crystallographic properties of CdTe thin films play a decisive role in governing their optical absorption, charge transport, and overall device performance, particularly in photovoltaic and optoelectronic applications. These properties are most commonly investigated using X-ray diffraction (XRD), which provides detailed insights into crystal structure, phase purity, lattice parameters, crystallite size, preferred orientation, microstrain, and defect density. CdTe thin films prepared by techniques such as chemical bath deposition, close-spaced sublimation, sputtering, and electrodeposition typically exhibit a polycrystalline nature with well-defined diffraction features. CdTe crystallizes predominantly in the cubic zinc blende structure [10-11], which is thermodynamically stable at room temperature. The XRD patterns of CdTe thin films generally show characteristic diffraction peaks corresponding to planes such as (111), (220), (311), (400), and (511). Among these, the (111) plane, appearing at $2\theta = 23.8°$, is usually the most intense, indicating a strong preferred orientation along the (111) direction. This preferred orientation arises due to the minimization of surface energy during film growth and is particularly beneficial for enhancing charge transport across the film.

The observed diffraction peaks are in close agreement with standard reference data reported in JCPDS card no. 15-0770, confirming the formation of single-phase cubic CdTe. The interplanar spacing corresponding to each diffraction peak is calculated using Bragg's law with Cu Kα radiation ($\lambda = 1.5406$ Å). For CdTe thin films, typical d-values are found to be approximately 3.73 Å for the (111) plane, 2.29 Å for the (220) plane, 1.95 Å for the (311) plane, and 1.62 Å for the (400) plane. These values are in excellent agreement with standard crystallographic data, confirming the high crystalline quality of the films.

For a cubic crystal system, the lattice parameter (a) is determined using the relation involving interplanar spacing and Miller indices. The calculated lattice constant for CdTe thin films typically lies in the range of 6.45–6.50 Å, which closely matches the standard bulk value of 6.48 Å. This close agreement indicates minimal lattice distortion and confirms that the deposited films possess good crystallinity. Slight variations in the lattice parameter may occur due to internal strain, defects, or deviations from ideal stoichiometry during film growth. The crystallite size of CdTe thin films is estimated using the Scherrer equation, which relates the broadening of diffraction peaks to the size of coherently diffracting domains. For typical full width at half maximum (FWHM) values in the range of 0.002–0.006 radians, the crystallite size is found to lie between 50 and 200 nm, indicating a nanocrystalline to microcrystalline structure. For instance, considering the (111) peak at $2\theta = 23.8°$ with a FWHM of about 0.003 radians, the crystallite size is estimated to be around 100–120 nm. It is generally observed that higher deposition temperatures or post-deposition annealing treatments promote grain growth, resulting in larger crystallite sizes and improved structural order. Microstrain (ε), which represents lattice distortion arising from defects, dislocations, and grain boundaries, is evaluated from peak broadening analysis. For CdTe thin films, the microstrain typically lies in the range of $(0.5–3) \times 10^{-3}$, indicating relatively low lattice distortion. Lower microstrain values are associated with improved crystal quality and better electronic properties. The dislocation density (δ), which is a measure of the number of defects in the crystal lattice, is calculated as the inverse square of the crystallite size. For CdTe thin films with crystallite sizes between 50 and 200 nm, the dislocation density is found to be of the order of 10^{13}–10^{14} lines/m². This relatively low defect density is advantageous for reducing charge carrier recombination and enhancing device efficiency. The crystallographic properties of CdTe thin films are strongly influenced by deposition conditions such as temperature, precursor concentration, and growth rate, as well as post-deposition treatments. Annealing treatment leads to sharper and more intense XRD peaks, increased grain size, reduced microstrain, and lower defect density. Film thickness also affects crystallographic behavior. Thicker films generally exhibit better crystallinity due to enhanced grain growth and reduced substrate influence, while thinner films may show broader peaks due to smaller grain sizes and higher defect density.

4. Optical Properties of CdTe Thin Films

The optical properties of CdTe thin films are of paramount importance for their application in photovoltaic devices, photodetectors, and other optoelectronic systems. Optical characterization is typically carried out using UV–Visible spectroscopy in the wavelength range of 300–1100 nm, enabling the evaluation of key parameters such as absorption coefficient, transmittance, reflectance, optical band gap, refractive index, and dielectric constants.

CdTe thin films exhibit strong optical absorption due to their direct band gap nature. The absorption coefficient (α) is typically greater than 10^5 cm^{-1} for photon energies above the band gap, indicating that CdTe can absorb a large fraction of incident light within a very short penetration depth. This high absorption coefficient allows CdTe films to absorb approximately 90–99% of incident solar radiation within a thickness of 1–2 μm, making them highly efficient absorber layers in thin film solar cells. The optical transmittance of CdTe thin films depends strongly on film thickness and wavelength. In general, CdTe films show low transmittance (10–40%) in the visible region due to strong absorption, while transmittance increases in the near-infrared region. Thinner films (200–500 nm) may exhibit relatively higher transmittance, whereas thicker films (1–2 µm) become nearly opaque in the visible range. The reflectance of CdTe thin films typically lies between 10% and 25%, depending on surface roughness and film quality. One of the most critical optical parameters is the optical band gap (Eg), which determines the energy threshold for electronic transitions. CdTe is a direct band gap semiconductor, and its band gap is commonly determined using the Tauc relation for direct allowed transitions. By plotting $(\alpha h\nu)^2$ versus photon energy ($h\nu$) and extrapolating the linear region to the energy axis, the band gap is obtained. For CdTe thin films, the band gap is typically found to be in the range of 1.44–1.50 eV, with the most commonly reported value being 1.45 eV at room temperature. Slight variations in band gap may occur due to quantum confinement effects in nanocrystalline films, strain, defects, or compositional variations [12-13].

The sharpness of the absorption edge provides insight into the structural quality of the films. A sharp absorption edge indicates good crystallinity and fewer defects, whereas a broadened edge suggests the presence of localized states within the band gap. These localized states are quantified using the Urbach energy (Eu), which represents the width of the tail of the absorption edge. For CdTe thin films, the Urbach energy typically lies in the range of 0.05–0.15 eV, indicating relatively low disorder compared to many other thin film materials. The refractive index (n) of CdTe thin films is another important optical parameter and is typically found to be in the range of 2.5–3.0 in the visible region. This relatively high refractive index indicates strong interaction between light and the material. The extinction coefficient (k), which represents the attenuation of light within the material, generally lies between 0.2 and 0.8, depending on wavelength and film thickness.

From these optical constants, the dielectric properties of CdTe thin films can be evaluated. The real part of the dielectric constant (ε_1), which represents the ability of the material to store optical energy, typically ranges from 6 to 10, while the imaginary part (ε_2), associated with energy loss, is higher in regions of strong absorption. The optical properties of CdTe thin films are strongly influenced by deposition conditions and post-deposition treatments. Annealing, particularly in the presence of $CdCl_2$ at temperatures between 300°C and 450°C, significantly improves crystallinity, reduces defect states, and sharpens the absorption edge. This result in more well-defined band gap values and improved optical performance. Film thickness also plays a critical role. As thickness increases, absorption increases and transmittance decreases, while the band gap may show slight variation due to changes in grain size and internal stress. Surface morphology and roughness further influence light scattering and overall optical response.

5. Electrical Properties of CdTe Thin Films

The electrical properties of CdTe thin films are of critical importance for their performance in photovoltaic and optoelectronic devices. These properties govern charge carrier generation, transport, and recombination processes, which directly influence device efficiency. CdTe thin films exhibit well-defined semiconducting behavior, and their electrical characteristics depend strongly on factors such as crystallinity, defect density, doping, grain size, and post-deposition treatments.

CdTe is a direct band gap semiconductor, and its intrinsic electrical conductivity is relatively low. The electrical conductivity (σ) of CdTe thin films typically lies in the range of 10^{-8} to 10^{-3} S/cm for as-deposited films. However, after suitable doping and annealing treatments, the conductivity can increase significantly to values as high as 10^{-2} to 10^{0} S/cm. The corresponding resistivity (ρ) usually ranges from 10^{3} to 10^{7} Ω·cm for intrinsic or lightly doped films, and can decrease to 10^{1}–10^{3} Ω·cm for highly doped or well-annealed films. The temperature dependence of electrical conductivity in CdTe thin films follows the Arrhenius-type behavior, indicating thermally activated conduction. The activation energy for CdTe thin films typically lies in the range of 0.1–0.6 eV, depending on the level of doping and defect states. Lower activation energies are associated with higher carrier concentrations and improved conductivity [14-15].

The conduction mechanism in CdTe thin films varies with temperature and structural quality. At higher temperatures, conduction is dominated by band conduction, where electrons and holes are thermally excited across the band gap. At lower temperatures or in films with higher defect density, conduction occurs through hopping mechanisms, such as nearest-neighbor hopping or variable range hopping, involving localized states within the band gap. CdTe thin films can exhibit both n-type and p-type conductivity [16], depending on the type of defects and dopants present. Intrinsically, CdTe tends to show p-type conductivity, which is primarily attributed to the presence of cadmium vacancies acting as acceptor states. These vacancies create hole carriers, making p-type conduction dominant. However, n-type conductivity can also be achieved through doping with donor impurities such as chlorine or indium, which introduce electron carriers. The carrier concentration in CdTe thin films typically lies in the range of 10^{13} to 10^{16} cm^{-3} for intrinsic films, and can increase to 10^{16}–10^{18} cm^{-3} with doping. The carrier mobility is relatively high compared to many other thin film semiconductors, with electron mobility ranging from 100 to 1100 $cm^2/V{\cdot}s$ and hole mobility between 40 and 100 $cm^2/V{\cdot}s$ [17-19]. However, in thin films, the effective mobility is often lower due to scattering at grain boundaries, defects, and impurities. Grain boundaries play a significant role in determining the electrical behavior of CdTe thin films. In polycrystalline films, grain boundaries can act as potential barriers, trapping centers, or recombination sites, which impede charge carrier transport. The barrier height at grain boundaries typically lies in the range of 0.1–0.3 eV, affecting carrier mobility and conductivity.

Post-deposition treatments such as annealing help to reduce these barriers by promoting grain growth and passivating defects.

Post-deposition treatments, particularly $CdCl_2$ annealing at temperatures between 300°C and 450°C, play a crucial role in enhancing the electrical properties of CdTe thin films. This treatment increases grain size, reduces defect density, improves carrier mobility, and enhances conductivity. It also activates dopants and improves junction properties in photovoltaic devices. Film thickness also influences electrical properties. Thicker films generally exhibit lower resistivity due to improved grain connectivity and reduced influence of substrate-induced defects, whereas thinner films may show higher resistivity due to increased scattering and defect density.

6. Photoelectrochemical (PEC) Properties of CdTe Thin Films

The PEC properties of CdTe thin films are of great importance for their application in solar energy conversion, photodetectors, and water-splitting devices. These properties describe the ability of CdTe to absorb light, generate charge carriers, and participate in electrochemical reactions at the semiconductor–electrolyte interface. Due to its direct band gap and high absorption coefficient, CdTe is highly efficient in harvesting solar radiation and converting it into electrical or chemical energy. When CdTe thin films are illuminated with light of energy equal to or greater than their band gap, photons are absorbed and excite electrons from the valence band to the conduction band, leaving behind holes. These photogenerated carriers are separated under the influence of the built-in electric field at the semiconductor–electrolyte junction. The separation efficiency depends on the quality of the film, crystallinity, and defect density. The electrons and holes then participate in reduction and oxidation reactions at the electrode surface, generating a measurable photocurrent. CdTe thin films generally exhibit p-type conductivity, and therefore, they typically act as photocathodes in PEC cells. Under illumination, electrons move toward the surface and participate in reduction reactions, while holes move toward the bulk or back contact. The PEC performance is usually studied using a three-electrode configuration consisting of the CdTe thin film as the working electrode, a platinum counter electrode, and a reference electrode such as Ag/AgCl, immersed in an electrolyte solution commonly Na_2S, Na_2SO_4, or polysulfide electrolytes. The photocurrent density (Jph) is one of the most important parameters used to evaluate PEC performance. For CdTe thin films, the photocurrent density typically lies in the range of 0.5 to 5mA/cm^2 under standard illumination conditions. Higher photocurrent values are associated with improved crystallinity, larger grain size, and reduced recombination losses. The dark current is usually very low, indicating good rectifying behavior and efficient charge separation.

The photovoltage generated in CdTe PEC systems typically ranges from 0.4 to 0.8 V, depending on the electrolyte and electrode configuration. The overall photoelectrochemical conversion efficiency is generally in the range of 2–8% [20-21], although higher efficiencies can be achieved with optimized device structures and surface treatments. The spectral response of CdTe thin films extends over a wide range of the solar spectrum, particularly from 400 nm to 850 nm, which corresponds to the visible region where solar irradiance is maximum. The incident photon-to-current efficiency or external quantum efficiency of CdTe thin films typically reaches values of 50–80% in the visible region, indicating efficient conversion of absorbed photons into charge carriers. The performance of CdTe in PEC applications is strongly influenced by surface states and recombination processes. Surface defects and trap states can act as recombination centers, reducing photocurrent and overall efficiency. To overcome this, surface modification techniques such as cadmium chloride treatment, chemical etching, or deposition of passivation layers are commonly employed. These treatments reduce surface recombination, improve charge separation, and enhance PEC performance.

The flat band potential, which is an important parameter in PEC studies, typically lies in the range of −0.5 to −0.8 V vs. Ag/AgCl for p-type CdTe. This parameter determines the band alignment and influences the efficiency of charge transfer at the interface. Stability is another important aspect of PEC performance. CdTe thin films generally exhibit moderate stability in aqueous electrolytes; however, long-term operation may lead to photocorrosion or surface degradation. The use of protective coatings or stable electrolytes can significantly enhance durability. Film thickness and morphology also affect PEC properties. Thicker films (1–3 µm) provide better light absorption but may suffer from increased recombination if carrier diffusion length is limited. Nanostructured or porous films offer higher surface area, which enhances electrochemical activity but may also increase defect density if not well controlled [22-23].

References

1. Tivanov, M., Zgliui, A., Gevorgyan, V., Korolik, O., Kolesov, E., Haroyan, K., Zhyhulin, D., & Gremenok, V. (2025). *Journal of Materials Science, 60*, 21250.
2. Koshelev, I., Volchkov, I., Podkur, P., Khairetdinova, D., Doludenko, I., & Kanevsky, V. (2024). *Crystallography Reports, 69*, 235.
3. Tuna, A., Gezgin, S., Mercimek, B., & Kiliç, H. (2026). *Optical and Quantum Electronics, 58*, 126.
4. Volchkov, I., Koshelev, I., Khairetdinova, D., Reznikova, P., Khmelenin, D., Butashin, A., & Kanevsky, V. (2026). *Thin Solid Films, 833*, 140835.
5. Tivanov, M., Gevorgyan, V., Zgliui, A., Haroyan, K., Gremenok, V., Zhyhulin, D., Dong, M., Hossain, M., Sayyed, M., Zubar, T., Tishkevich, D., & Trukhanov, A. (2025). *Journal of Materials Science: Materials in Electronics, 36*, 699.
6. Bandara, K., Lakmal, A., Perera, H., Dissanayake, M., Seneviratne, V., Delclos, T., Dassanayake, B., & Das, G. (2026). *Applied Surface Science Advances, 32*, 100958.
7. Elmourabit, F., Essakali, Y., & Dlimi, S. (2026). *Hybrid Advances, 12*, 100582.

8. Jena, I., & Singh, U. (2025). *Journal of Materials Science: Materials in Electronics, 36*, 132.
9. Chate, P., & Sathe, D. (2019). *Metal chalcogenide nanostructures: Characteristics and synthesis*. OMICS International.
10. Parida, P., Patra, J., Singh, V., Rahaman, A., Patel, P., Mahapatra, S., Sahoo, S., & Verma, V. (2025). *Physica B, 715*, 417567.
11. Shinde, A., Salunke, S., Patil, R., Deshmukh, R., & Joshi, Y. (2026). *Materials Letters, 409*, 140234.
12. Surucu, O., Surucu, G., Gasanly, N., Parlak, M., & Isik, M. (2024). *Materials Letters, 373*, 137078.
13. Sharmin, A., Mahmood, S., Sultana, M., Aziz, S., Shaikh, M., & Bashar, M. (2023). *Journal of Materials Science: Materials in Electronics, 34*, 344.
14. Verma, L., & Khare, A. (2024). *Emergent Materials, 7*, 2065.
15. Kachhia, Z., Chaki, S., Patel, S., Tailor, J., Parekh, Z. R., & Deshpande, M. (2024). *Next Materials, 3*, 100152.
16. Oluyamo, S., Faremi, A., Olusola, O., & Odusote, Y. (2021). *Materials Today: Proceedings, 38*, 558.
17. Segall, B., Lorenz, M., & Halsted, R. (1963). *Physical Review, 129*, 2471.
18. Gumaste, M., & Kulkarni, G. (2021). *Materials Today: Proceedings, 45*, 3927.
19. Gutowski, J., Sebald, K., & Voss, T. (2008). CdTe: Mobility. In U. Roessler (Ed.), *Semiconductors – CdTe: Mobility*. Springer Materials.
20. Santos, I., Pérez-Orozco, A., Liña-Martínez, D., Courel, M., Meza-Avendaño, C., Borrego-Pérez, J., Pérez, L., & Laroze, D. (2023). *Nanomaterials, 13*, 1335.
21. Kirk, A. (2024). *Journal of Alloys Metallurgical Systems, 5*, 100050.
22. Kondratenko, S., & Mykytiuk, A. (2024). *Optical Materials, 157*, 116061.
23. Kosyachenko, L. (2006). *Semiconductors, 40*, 710.

RECENT ADVANCES IN HYDROGEN ENERGY GENERATION: MECHANISMS, MATERIALS, AND STRATEGIES FOR EFFICIENT WATER SPLITTING

Dilip Saikia

Department of Physics,

Sialpathar College, Sialpathar, Dhemaji, Assam 787059

Corresponding author E-mail: dilip.saikia202@gmail.com

Abstract

The progress in human lifestyle comes at the cost of rising energy demand, and it mainly relies on fossil-based energy. The use of fossil-based energy largely contributes to the generation of greenhouse gases, resulting in adverse environmental effects such as global warming. Reducing fossil fuel consumption can help mitigate environmental pollution, but the resulting consequences may significantly affect human lifestyles. Hydrogen can be considered as an ideal candidate for an energy solution with a high calorific value of about 142 MJ/Kg (almost 3 times higher than that of diesel/gasoline). Hydrogen can be derived from purely renewable resources and can be utilised as a truly sustainable source of energy. Here we will discuss the generation of hydrogen by splitting water.

1. Introduction

The rapid progress of industrialization and the ever-increasing global energy demand have resulted in a heavy reliance on fossil fuels that still control the world energy market. But their excessive use has created huge environmental problems, especially the emission of greenhouse gases like CO_2, which causes global warming and climate change[1]. Growing environmental awareness, strict regulatory frameworks, and the pressing necessity for energy security and carbon neutrality have accelerated the shift toward renewable and sustainable energy systems. Future projections suggest that global energy requirements and CO_2 emissions will continue to increase, emphasizing the urgent need for cleaner energy solutions[2]. Renewable energy sources such as solar, wind, and hydropower offer environmentally friendly alternatives but suffer from intermittency and variability, posing challenges for a reliable energy supply. In this context, hydrogen has emerged as a promising clean energy alternative due to its high energy density (~142 MJ/kg), clean combustion, and compatibility with renewable systems[5]. Hydrogen fuel cells enable efficient energy conversion with water as the only by-product, and hydrogen is expected to contribute significantly to future energy systems, potentially accounting for ~12% of global energy consumption by 2050[3].

Hydrogen can be produced via several methods, including Steam Methane Reforming (SMR), coal gasification, biomass conversion, and water splitting. While SMR is cost-effective, it is carbon-intensive. In contrast, green hydrogen produced via electrochemical, photocatalytic, and photoelectrochemical water splitting offers a sustainable alternative, though challenges such as high cost, low efficiency, and material limitations persist[4]. This chapter focuses on recent advances in hydrogen production pathways, mechanisms, and materials for sustainable energy applications.

2. Hydrogen Production Methods

Hydrogen is broadly recognized as a promising source of clean and green energy due to its high calorific value of energy density and environmentally friendly nature. However, its sustainability greatly depends on the production process employed. Currently, most of the industrial hydrogen is generated using fossil fuel-based processes, resulting in raised environmental issues. With the global transition toward low-carbon energy systems, renewable-driven hydrogen production methods are gaining increasing importance. Conventional techniques such as steam methane reforming, coal and biomass gasification, and water electrolysis remain widely used. Hydrogen can be produced from both renewable and non-renewable resources, with ongoing advancements improving efficiency and yield. Each method has specific advantages and limitations, motivating continued research to enhance economic and environmental viability. Fig. 1 presents fossil fuel-based pathways[5], while Fig. 2 highlights key water-splitting approaches: electrolysis, photoelectrochemical (photolysis), and thermochemical (thermolysis) methods.

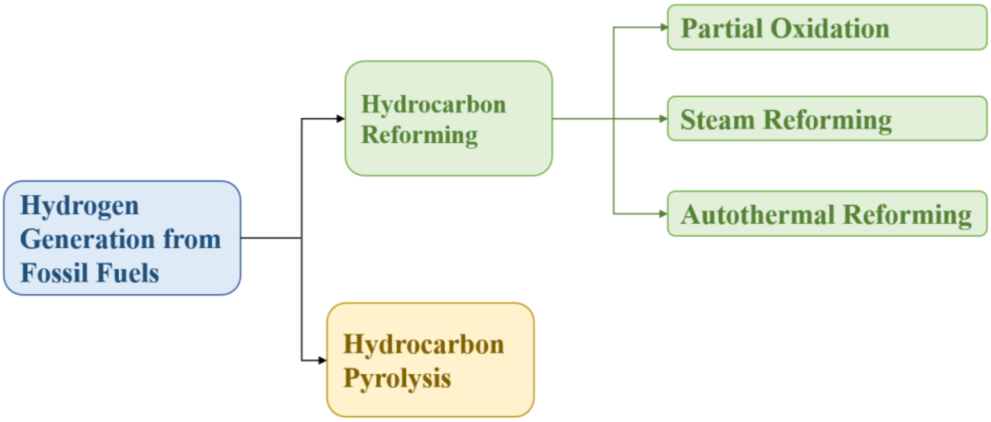


Figure 1: Hydrogen production methods from fossil fuels

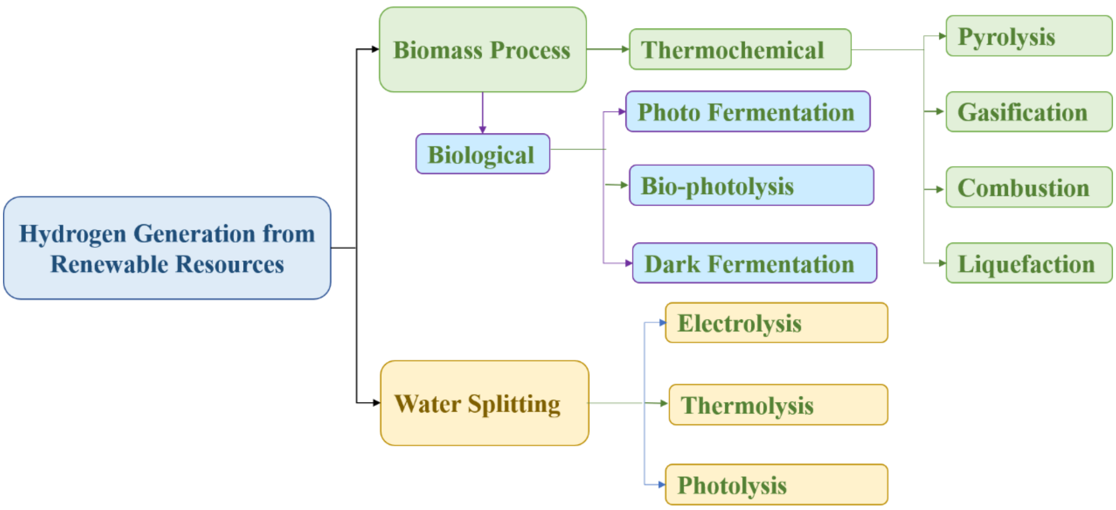


Figure 2: Hydrogen production methods from renewable resources

3. Water Splitting

Water is an abundant resource for sustainable hydrogen production via electrolysis, thermolysis, and photo-electrolysis. Among these, electrolysis is the most promising and eco-friendly, especially when powered by renewable energy. The two main technologies are alkaline electrolysis (AE) and proton exchange membrane (PEM) electrolysis[6]. AE uses nickel-based electrodes and alkaline electrolytes, while PEM employs proton-conducting membranes like Nafion[7]. Advanced catalysts such as Pt, Ir, Ru, Ni, and Co enhance performance[8]. Ongoing improvements in catalysts and materials are driving efficient and cost-effective hydrogen production technologies.

3.1. Thermodynamics of water splitting

Water splitting is an uphill thermodynamic process requiring external energy to decompose water into hydrogen and oxygen:

$$2H_2O(l) \rightarrow 2H_2(g) + O_2(g)$$

The Gibbs free energy change ($\Delta G^\circ \approx 237.2$ kJ mol^{-1}) corresponds to a minimum potential of 1.23 V, though a higher voltage is needed due to overpotential[9]. The process involves hydrogen evolution reaction (HER) at the cathode and oxygen evolution reaction (OER) at the anode. HER is kinetically favorable, whereas OER is sluggish due to multi-electron transfer. Efficiency depends on catalysts, electrolyte, pH, and temperature, making catalyst development essential for improved performance.

3.2. Hydrogen Evolution Reaction (HER)

The Hydrogen Evolution Reaction (HER) is a key electrochemical process for hydrogen production, involving the reduction of protons or water at the cathode. It proceeds via three steps: Volmer (adsorption), Heyrovsky (electrochemical desorption), and Tafel (recombination)[10].

In acidic media[11] it can be written as:

$$H^+ + e^- \rightarrow H^{ab}$$

where H^{ab} represents an adsorbed hydrogen atom on the catalyst surface. This is followed by either the Heyrovsky reaction (electrochemical desorption):

$$H^{ab} + H^+ + e^- \rightarrow H_2$$

or the Tafel reaction (chemical recombination):

$$H^{ab} + H^{ab} \rightarrow H_2$$

In alkaline media, water dissociation replaces proton reduction:

$$H_2O + e^- \rightarrow H^{ab} + OH^-$$

HER kinetics depend on catalyst properties, surface structure, pH, and overpotential[12]. Efficient catalysts exhibit near-zero hydrogen adsorption energy ($\Delta G\_H^* \approx 0$). While Pt shows excellent activity, its high cost limits use, prompting research into alternatives like transition metal sulfides, phosphides, and carbides. HER remains vital for sustainable hydrogen production.

3.3. Oxygen Evolution Reaction (OER)

The Oxygen Evolution Reaction (OER) is the primary kinetic bottleneck in water splitting and related oxygen evolution processes. It governs the overall efficiency of hydrogen production and is fundamental to technologies such as solar fuels and metal–air batteries[13]. OER involves a complex four-electron/four-proton transfer, making it inherently sluggish and requiring high overpotentials.

Mechanistically, OER proceeds via pathways such as the adsorbate evolution mechanism (AEM) and lattice oxygen-mediated mechanism (LOM), which guide catalyst design. Transition metal-based materials—particularly oxides and hydroxides of Ni, Co, and Fe—have emerged as efficient and cost-effective alternatives to noble metals. Although RuO_2 and IrO_2 exhibit excellent activity, their scarcity limits practical use.

Earth-abundant materials, including manganese oxides, offer good stability under OER conditions. Strategies such as doping, defect engineering, heterostructure formation, and nano structuring enhance catalytic activity by optimizing intermediate adsorption energies[14]. Advanced approaches like interface engineering and in situ characterization provide insights into active sites and catalyst dynamics. These developments are crucial for designing efficient, stable, and scalable OER catalysts for sustainable hydrogen production.

4. Recent Advances in Water Splitting Technologies

Hydrogen (H_2) has emerged as a promising clean energy carrier due to its high gravimetric energy density (~120 MJ kg^{-1}) and environmentally benign nature, producing only water upon combustion[15]. It offers a sustainable alternative to fossil fuels and can be generated through diverse pathways using renewable and non-renewable resources. Among these, water splitting contributes ~4–5% of global hydrogen production and is increasingly recognized as a key route for green hydrogen generation[16].

Photocatalytic and photoelectrochemical (PEC) water splitting utilize solar energy to convert water into hydrogen and oxygen. In PEC systems, light absorption generates electron–hole pairs, followed by charge separation and migration to the surface, where redox reactions occur. The half-reactions are[17]:

$$2H^+ + 2e^- \rightarrow H_2,\ E^0_{red} = 0V$$

$$2H_2O + 4h^+ \rightarrow 4H^+ + O_2,\ E^0_{ox} = 1.23V$$

Efficient operation requires the conduction band (CB) to be more negative than 0 V and the valence band (VB) more positive than 1.23 V vs normal hydrogen electrode (NHE)[18]. Additionally, semiconductors must exhibit strong light absorption, efficient charge transport, and high stability.

Recent advances in photocatalysis focus on bandgap engineering, heterojunction design, and cocatalyst integration to enhance light absorption and reduce recombination. Electrochemical water splitting using alkaline and proton exchange membrane (PEM) electrolyzers has progressed through the development of cost-effective electrocatalysts and improved electrode architectures. PEC systems further integrate light absorption and electrolysis, achieving higher theoretical efficiencies.

Solid oxide electrolysis cells (SOECs) operate at high temperatures, improving efficiency through thermal energy utilization, with perovskite and Ni-based materials enhancing performance. Biological hydrogen production using microorganisms offers an eco-friendly alternative, though still in early stages. Overall, advances in materials, catalysis, and system integration are driving scalable hydrogen production technologies.

5. Challenges and Future Perspectives

Despite notable progress in water-splitting technologies, several challenges hinder large-scale hydrogen production. A key issue is balancing efficiency and durability, as high-performance catalysts and photoelectrodes often degrade under prolonged operation due to photocorrosion, poisoning, and structural instability[19]. In photocatalytic and photoelectrochemical systems, solar-to-hydrogen efficiency is limited by rapid charge recombination and poor visible-light absorption. Similarly, although noble metal catalysts exhibit high activity in electrochemical systems, their scarcity and cost restrict practical application, while earth-abundant alternatives still face efficiency and stability challenges.

Economic feasibility remains constrained by expensive materials, complex fabrication, and energy-intensive processes, particularly in solid oxide electrolysis systems. Scaling up also introduces issues such as reactor design, mass transport, and system integration, alongside hydrogen storage and distribution challenges.

Future research should focus on designing durable, cost-effective materials using advanced characterization and computational tools. Strategies such as heterostructure engineering, defect control, and interfacial optimization can enhance performance. Integration of AI, scalable reactor design, and renewable energy coupling will be crucial for sustainable hydrogen production[20].

Conclusion

In summary, this chapter emphasizes the significant developments achieved in the generation of hydrogen energy through a variety of water-splitting technologies, such as photocatalytic, electrochemical, photoelectrochemical, solid oxide, and biological approaches. Substantial progress in the fundamental understanding of HER and OER mechanisms, coupled with innovations in functional materials and catalyst engineering, has led to prominent enhancements in hydrogen production efficiency. Nevertheless, significant challenges remain, basically in attaining a good balance between efficiency and durability under prolonged operation, minimizing dependence on expensive noble metals, and addressing issues of scalability and infrastructure development. The transition from laboratory-scale developments to industrial implementation continues to pose a significant challenge. Future research should emphasize the rational design of robust, earth-abundant materials, deeper insights into underlying mechanisms, and the incorporation of emerging techniques such as artificial intelligence/machine learning to accelerate innovation and performance optimization. Ultimately, sustained interdisciplinary collaboration and system-level innovations will be decisive for achieving economically viable and sustainable hydrogen production in support of a clean and green energy of the future.

References

1. Ibekwe, K. I., Etukudoh, E. A., Nwokediegwu, Z. Q. S., Umoh, A. A., Adefemi, A., & Ilojianya, V. I. (2024). Energy security in the global context: A comprehensive review of geopolitical dynamics and policies. *Engineering Science & Technology Journal*.
2. Zhao, G., Rui, K., Dou, S. X., & Sun, W. (2018). Heterostructures for electrochemical hydrogen evolution reaction: A review. *Advanced Functional Materials, 28*(43), 1803291.
3. Schlapbach, L., & Züttel, A. (2001). Hydrogen-storage materials for mobile applications. *Nature, 414*(6861), 353–358.
4. Youssef, M. M. M., Maghraby, A., Abdel-Akher, M., & Hamdan, I. (2026). Recent advances in hydrogen production technologies: Environmental and economic perspectives. *Journal of Engineering and Applied Science, 73*(1).
5. Nikolaidis, P., & Poullikkas, A. (2017). A comparative overview of hydrogen production processes. *Renewable and Sustainable Energy Reviews, 67*, 597–611.
6. Hwang, J., Maharjan, K., & Cho, H. (2023). A review of hydrogen utilization in power generation and transportation sectors: Achievements and future challenges. *International Journal of Hydrogen Energy, 48*(74), 28629–28648.
7. Malik, F. R., Yuan, H.-B., Moran, J. C., & Tippayawong, N. (2023). Overview of hydrogen production technologies for fuel cell utilization. *Engineering Science and Technology, an International Journal, 43*, 101452.
8. Zhen, Z., Wang, Y., Wang, Y., Wang, X., Ou, X., & Zhou, S. (2023). Hydrogen production paths in China based on learning curve and discrete choice model. *Journal of Cleaner Production, 415*, 137848.
9. Liu, F., Shi, C., Guo, X., He, Z., Pan, L., Huang, Z.-F., Zhang, X., & Zou, J.-J. (2022). Rational design of better hydrogen evolution electrocatalysts for water splitting: A review. *Advanced Science, 9*(18), 2200307.
10. Conway, B. E., & Tilak, B. V. (2002). Interfacial processes involving electrocatalytic evolution and oxidation of H_2, and the role of chemisorbed H. *Electrochimica Acta, 47*(22), 3571–3594.
11. Kahyarian, A., Singer, M., & Nesic, S. (2016). Modeling of uniform CO_2 corrosion of mild steel in gas transportation systems: A review. *Journal of Natural Gas Science and Engineering, 29*, 530–549.

12. Juodkazis, K., Juodkazytė, J., Grigucevičienė, A., & Juodkazis, S. (2011). Hydrogen species within the metals: Role of molecular hydrogen ion H_2^+. *Applied Surface Science, 258*(2), 743–747.
13. Tahir, M., Pan, L., Idrees, F., Zhang, X., Wang, L., Zou, J.-J., & Wang, Z. L. (2017). Electrocatalytic oxygen evolution reaction for energy conversion and storage: A comprehensive review. *Nano Energy, 37*, 136–157.
14. Shen, Y., Luo, Y., Li, J., Liu, Z., Liao, Y., Shi, B., Cheng, L., Li, Z., Yang, X., & Jiao, Z. (2026). Engineering carbon nanofiber-supported NiCo/$CoNi_2S_4$ Mott-Schottky heterostructure with robust interfacial electric field for boosting oxygen evolution reaction kinetics. *Journal of Colloid and Interface Science, 719*, 140649.
15. Dawood, F., Anda, M., & Shafiullah, G. M. (2020). Hydrogen production for energy: An overview. *International Journal of Hydrogen Energy, 45*(7), 3847–3869.
16. Minggu, L. J., Daud, W. R. W., & Kassim, M. B. (2010). An overview of photocells and photoreactors for photoelectrochemical water splitting. *International Journal of Hydrogen Energy, 35*(11), 5233–5244.
17. Chen, S., Liu, T., Zheng, Z., Ishaq, M., Liang, G., Fan, P., Chen, T., & Tang, J. (2022). Recent progress and perspectives on Sb_2Se_3-based photocathodes for solar hydrogen production via photoelectrochemical water splitting. *Journal of Energy Chemistry, 67*, 508–523.
18. Rodríguez-Lazcano, Y., Peña, Y., Nair, M. T. S., & Nair, P. K. (2005). Polycrystalline thin films of antimony selenide via chemical bath deposition and post deposition treatments. *Thin Solid Films, 493*(1), 77–82.
19. Shehzad, B., Latif, S., Wahab, A., Shafique, M. U., & Ch, A. R. (2026). Progress and challenges in alkaline water electrolysis for sustainable hydrogen production. *Discover Chemistry, 3*(1).
20. Otyepka, M., Pykal, M., & Otyepka, M. (2025). Advancing materials discovery through artificial intelligence. *Applied Materials Today, 47*, 102981.

Applied Research in Science, Technology and Engineering Volume I

(ISBN: 978-93-47587-00-9)

About Editors

Dr. Prerakkumar A. Patel earned his Ph.D. in Mathematics from Gujarat University and qualified CSIR-NET with AIR-31. He is presently working as Assistant Professor at Gujarat Technological University–Institute of Technology and Research, Mevad, and possesses more than nine years of teaching experience. His research interests include Differential Equations, Fractional Calculus, Wavelet Methods, and Graph Theory. Dr. Patel serves as Principal Investigator of a funded research project on mathematical modeling of solar air heaters and holds three registered national design patents. He has published several research papers in reputed international journals and authored or edited books in Mathematics Education and Reasoning. He has also contributed a book chapter on sustainability in academia and actively guides research scholars. In 2015, he received the Minaxi Lalit Science Award from the Gujarat Science Academy for excellence in Mathematics research and education.

Dr. Narayan Dattatraya Totewad completed his B.Sc. from N.E.S. Science College, Nanded, and earned his M.Sc. in Microbiology and Ph.D. in Biotechnology from S.R.T.M. University, Nanded He received the UGC Rajiv Gandhi National JRF and SRF during 2007–2012 and completed a PG Diploma in Bio-nanotechnology with distinction from B.K. Birla College, Kalyan, in 2020. Presently, he serves as Assistant Professor in Microbiology at B.K. Birla College (Autonomous), Kalyan, and is a recognized PG guide under the University of Mumbai He possesses 11 years of teaching and 18 years of research experience, with 17 research publications and 28 granted patents He has edited 12 books, authored 15 book chapters, and received the Excellence and Innovation Award in Microbiology in 2020. His research interests include Microbiology, Bio-nanotechnology, Environmental Science, and Biomedical Sciences.

Dr. Narayan Gaonkar is presently serving as Assistant Professor of Physics at University College of Science, a constituent college of Tumkur University, Tumkur, Karnataka. He obtained his Master's degree in Physics from Karnatak University, Dharwad, in 2009 with gold medals and qualified CSIR-NET in 2010. He earned his Ph.D. from Tumkur University in 2023. His research interests are mainly focused on theoretical physics and related interdisciplinary studies. Dr. Gaonkar has published 16 research papers in reputed journals and contributed a book chapter in his field of specialization. He has actively participated in more than 20 national and international conferences, presenting his research findings. In addition, he has jointly authored three undergraduate physics textbooks, contributing significantly to physics education and academic development. He continues to engage in teaching, research, and scholarly activities in Physics.

Dr. Pramod Machindra Dhakane is working as Associate Professor and Head of the Department of Mathematics at S.B.E.S. College of Science, Chhatrapati Sambhajinagar, Maharashtra, India. He obtained his Ph.D. in Mathematics from Dr. Babasaheb Ambedkar Marathwada University, Aurangabad. His research interests include Integral Equations and Differential Equations. Dr. Dhakane possesses 26 years of teaching experience at the undergraduate level and two years at the postgraduate level. He has published seven research papers in national and international journals and presented eight research papers at various conferences and seminars. He has also authored four books in the field of Mathematics. In addition, he has actively participated in 15 national and international research conferences, contributing significantly to academic and research activities in Mathematics education and allied areas.

Bhumi Publishing

Bhumi Publishing,
Kolhapur, Maharashtra, India

bhumipublishing@gmail.com

www.bhumipublishing.com

ISBN 978-93-475-8700-9

9 789347 587009